



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 117804

To: Dave Nguyen
Location: rem/2d31/2c18
Art Unit: 1632
Monday, March 29, 2004

Case Serial Number: 09/869508

From: Beverly Shears
Location: Remsen Bldg.
RM 1A54
Phone: 571-272-2528

beverly.shears@uspto.gov

Search Notes

SEARCH REQUEST FORM

Requestor's
Name: _____

Serial
Number: _____

Date: _____

Phone: _____

Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 03-29-04
 Searcher: Beverly C 2528
 Terminal time: 20
 Elapsed time: _____
 CPU time: _____
 Total time: 23
 Number of Searches: _____
 Number of Databases: 1

Search Site

____ STIC
 ____ CM-1
 ____ Pre-S

Type of Search

____ N.A. Sequence
 ____ A.A. Sequence
 ____ Structure
 ____ Bibliographic

Vendors

____ IG
 ____ STN
 ____ Dialog
 ____ APS
 ____ Geninfo
 ____ SDC
 ____ DARC/Questel
☒ Other CGN

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 12:40:25 ; Search time 6740 Seconds
(without alignments)
10919.358 Million cell updates/sec

Title: US-09-869-508-1_COPY_453_2150

Perfect score: 1698

Sequence: 1 agaacccatggtggccaa.....caccocggaggagccgcag 1698

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.ov.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.ste.*

28: em.un.*

29: em.vi.*

30: em.htg_hum.*

31: em.htg_inv.*

32: em.htg_other.*

33: em.htg_mus.*

34: em.htg_pln.*

35: em.htg_rod.*

36: em.htg_man.*

37: em.htg_vrt.*

38: em.sy.*

39: em.htgo_hum.*

40: em.htgo_mus.*

41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1698	100.0	2550	9	HSYB1	X96666 H.sapiens Y
2	1672.4	98.5	161835	9	AC098484	AC098484 Homo sapi
3	456.6	26.9	659	9	HUMY292E04	AF086102 Homo sapi
C 4	186.8	11.0	162454	10	EX470216	EX470216 Mouse DNA
C 5	186.8	11.0	194020	2	AC0101988	AC0101988 Mus muscu
C 6	186.2	11.0	225308	2	AC098918	AC098918 Rattus no
7	111	6.5	178	6	AX897942	AX897942 Sequence
8	111	6.5	178	6	BD033475	BD033475 Sequence
9	107.4	6.3	1548	6	AX014868	AX014868 Sequence
10	107.4	6.3	1548	6	BD033713	BD033713 Human nuc
11	101.4	6.0	144902	2	AC016114	AC016114 Homo sapi
12	101.4	6.0	162144	9	CNS06C85	AL391261 Human chr
C 13	101.4	6.0	184656	9	CNS05TDB	AL356021 Human chr
C 14	96.6	5.7	189367	9	AL162733	AL162733 Human DNA
C 15	95	5.6	99886	9	AL135841	AL135841 Human DNA
16	89.4	5.3	395	6	AX321258	AX321258 Sequence
17	89	5.2	1481	6	AR117695	AR117695 Sequence
18	89	5.2	1481	6	AX409453	AX409453 Sequence
19	89	5.2	1481	6	AX498410	AX498410 Sequence
20	89	5.2	1481	6	AX840171	AX840171 Sequence
21	89	5.2	1481	9	HUMYB1A	J03827 Y-box bindi
22	87.4	5.1	2168	9	HUMPSDBPB	L37516 Homo sapien
23	86.6	5.1	162996	2	AC051655	AC051655 Homo sapi
24	85.2	5.0	410	6	AX321869	AX321869 Sequence
25	84	4.9	1521	9	BC038384	BC038384 Homo sapi
26	83	4.9	3073	6	AR083654	AR083654 Sequence
27	83	4.9	3073	6	122491	122491 Sequence 31
28	82	4.8	917	9	BC000064	BC000064 Homo sapi
29	81.4	4.8	1468	9	HUMENABP	M83234 Human nucle
30	81.4	4.8	1474	9	HUMENABP	M85234 Human nucle
C 31	80.8	4.8	16805	2	AC147392	AC147392 Otolenur
32	80.6	4.7	142184	2	AC037428	AC037428 Homo sapi
33	80.6	4.7	143604	9	AC093610	AC093610 Homo sapi
34	80.6	4.7	310021	2	AC092883	AC092883 Homo sapi
C 35	79	4.7	77948	2	AC018982	AC018982 Homo sapi
36	78.8	4.6	381	6	BD275806	BD275806 COMPOUNDS
37	78.8	4.6	381	6	AR220591	AR220591 Sequence
38	78.8	4.6	381	6	AR255585	AR255585 Sequence
39	78.8	4.6	381	6	AR281155	AR281155 Sequence
40	78.8	4.6	381	6	AX365822	AX365822 Sequence
C 41	77.8	4.6	242756	2	AC093965	AC093965 Rattus no
42	77.4	4.6	545	6	AX321285	AX321285 Sequence
43	77	4.5	1554	9	BC010430	BC010430 Homo sapi
44	76.8	4.5	142565	10	AL592545	AL592545 Mouse DNA
45	76.8	4.5	197658	2	AL646094	AL646094 Mus muscu

ALIGNMENTS

RESULT 1	HSYB1	H.sapiens YB-1 gene promoter region.	2550 bp	DNA	linear	PRI 08-JUL-1996
LOCUS	HSYB1	H.sapiens YB-1 gene promoter region.	2550 bp	DNA	linear	PRI 08-JUL-1996
DEFINITION	X96666	X96666				
ACCESSION	X96666	X96666				
VERSION	X96666.1	GI:1403348				
KEYWORDS	promoter region; Y box binding protein; YB-1 gene.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	1					
AUTHORS	Makino, Y., Ohga, T., Toh, S., Koike, K., Okumura, K., Wada, M., Kuwano, M. and Kohno, K.					
TITLE	Structural and functional analysis of the human Y-box binding					

protein (YB-1) gene promoter
Nucleic Acids Res. 24 (10), 1873-1878 (1996)
MEDLINE 96296173
PUBMED 8657568
REFERENCE 2 (bases 1 to 2550)
AUTHORS Yoshinari M.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-1996) Yoshinari M., Department of Biochemistry,
Kyushu University School of Medicine, Maidashi, Fukuoka, 812-82,
JAPAN

FEATURES
source
1..2550
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
gene
1..2411
/gene="YB-1"
promoter
1..1855
/gene="YB-1"
exon
1856..2352
/gene="YB-1"
/number=1
repeat_region
1904..1914
/rpt_type="INVERTED"
protein_bind
1957..1962
/gene="YB-1"
/bound_moiety="SP-1"
protein_bind
2001..2006
/gene="YB-1"
protein_bind
2024..2029
/bound_moiety="SP-1"
protein_bind
2033..2038
/gene="YB-1"
repeat_region
2034..2044
/rpt_type="INVERTED"
protein_bind
2100..2105
/gene="YB-1"
CDS
2187..2411
/gene="YB-1"
/codon_start=1
/protein_id="CAA65446.1"
/db_xref="GI:1403349"
/translation="MSSEATQPPAAPPALPSAADTKPGCTGSGAGSGPGGLTS
APAGGDKRVIGSDROGWMGPRAAQRNR"

ORIGIN
Query Match 100.0%; Score 1698; DB 9; Length 2550;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGAACCATATGGTCGGCAAGCTTATATTATTTGCTCTCACAGAAAAGTT 60
Db 453 AGAACCATATGGTCGGCAAGCTTATATTATTTGCTCTCACAGAAAAGTT 512
Qy 61 CCTACCCCTTGCTCTAAGCAATTGGTGATGATGCGCTACAGTAGTGACAGTAGAATGAA 120
Db 513 CCTACCCCTTGCTCTAAGCAATTGGTGATGATGCGCTACAGTAGTGACAGTAGAATGAA 572
Qy 121 AAGACTCTGAATGACATTTAGAGGTTTAGGACTGACTCAAGAACCCCTTTAGGAGGTG 180
Db 573 AAGACTCTGAATGACATTTAGAGGTTTAGGACTGACTCAAGAACCCCTTTAGGAGGTG 632
Qy 181 GAACCTCAGGCCCTAGACGGCAATGGTAGGGGTAAAGCAATACCTGATTAAGCTCTTAAC 240
Db 533 GAACCTCAGGCCCTAGACGGCAATGGTAGGGGTAAAGCAATACCTGATTAAGCTCTTAAC 692
Qy 241 GCCTTCTCTACTCCATTCCTTCCCTCTGCAACCACTTCTGCAACCACTTCTGCAACCACTT 300

ORIGIN		/note="Single subclone region"	
Query Match		98.5%; Score 1672.4; DB 9; Length 161835;	
Best Local Similarity		99.8%; Pred. No. 0;	
Matches 1695; Conservative		0; Mismatches 1; Indels 2; Gaps 2;	
QY		1 AGAACCATATGTTGGTGGCCAAAGCTTATATTAATTAATTTGGCTCTCACAGAAAAGGTT 60	
Db		28812 AGAACCATATGTTGGTGGCCAAAGCTTATATTAATTAATTTGGCTCTCACAGAAAAGGTT 28871	
QY		61 CCTACCCCTTGGCTCTAGCAATTTGGTGATGATGCCCTACAGTAGTGACAGTAGAGATGAA 120	
Db		28972 CCTACCCCTTGGCTCTAGCAATTTGGTGATGATGCCCTACAGTAGTGACAGTAGAGATGAA 28931	
QY		121 AAGACTCTGAATTCACATTTAGAGGTTTAGGACTGACTCAAGAACGCCCTTTAGGAGGTG 180	
Db		28932 AAGACTCTGAATTCACATTTAGAGGTTTAGGACTGACTCAAGAACGCCCTTTAGGAGGTG 28991	
QY		181 GAACTCACAGCCCTAGACGGCATTTGGTAGGGTAAAGGAATCTGCAATTAAGCTCTTAACCT 240	
Db		28992 GAACTCACAGCCCTAGACGGCATTTGGTAGGGTAAAGGAATCTGCAATTAAGCTCTTAACCT 29051	
QY		241 GCTTTCTCTACTTCCATTCCTTGGCCCTCTGCAACCCATTTCTCCACTCCGACGCCATTTT 300	
Db		29052 GCTTTCTCTACTTCCATTCCTTGGCCCTCTGCAACCCATTTCTCCACTCCGACGCCATTTT 29111	
QY		301 TAAAGAGATGCCCTCCCTACTTATGACTCTAAATTTGCTTCTTCTCTCTCTCTCTCTCAG 360	
Db		29112 TAAAGAGATGCCCTCCCTACTTATGACTCTAAATTTGCTTCTTCTCTCTCTCTCTCTCAG 29171	
QY		361 GATATATTTCCAAATTAATATACCTAAGTCACTGCGCCACCTCTGCAACCCCAATGTCAAT 420	
Db		29172 GATATATTTCCAAATTAATATACCTAAGTCACTGCGCCACCTCTGCAACCCCAATGTCAAT 29231	
QY		421 TCGAGTCTTACTGAACTTACTGCAATTTCCGAGATCTCACCTCTCTCGCCTGTAC 480	
Db		29232 TCGAGTCTTACTGAACTTACTGCAATTTCCGAGATCTCACCTCTCTCGCCTGTAC 29291	
QY		481 CCTGTGGCGGAAAGTCAAGCCCTCCACTTCTCCCTGCTTCCACTCCCAAAATTAATTCGT 540	
Db		29292 CCTGTGGCGGAAAGTCAAGCCCTCCACTTCTCCCTGCTTCCACTCCCAAAATTAATTCGT 29351	
QY		541 GGTTTTGAGCTCTGGAGTATTTACCGTGTGGCTGTGTTTAAATTTCTGCCTCCATCAGAA 600	
Db		29352 GGTTTTGAGCTCTGGAGTATTTACCGTGTGGCTGTGTTTAAATTTCTGCCTCCATCAGAA 29411	
QY		601 GCGAGAACTGACTCGGCACTATTCCATCCCGAGCGATAGTAGAGCTTAAAGAAAGAA 660	
Db		29412 GCGAGAACTGACTCGGCACTATTCCATCCCGAGCGATAGTAGAGCTTAAAGAAAGAA 29471	
QY		661 CGAAGAAAGTGGGTGGGAGGACTTCAGTAACATCAGGTGGCAGCCTCAATTTTATCGTT 720	
Db		29472 CGAAGAAAGTGGGTGGGAGGACTTCAGTAACATCAGGTGGCAGCCTCAATTTTATCGTT 29530	
QY		721 TGTGAAACGTGGATAGTAATCCCTCTATCAGTGGCTGTGTCAGGAATAAGTGAAGAAA 780	
Db		29531 TGTGAAACGTGGATAGTAATCCCTCTATCAGTGGCTGTGTCAGGAATAAGTGAAGAAA 29590	
QY		781 CAAAACAGGCTAGCTTGTTCATTAATGTGAGTTGAATTAATCTGATTTGTGTCAGTA 840	
Db		29591 CAAAACAGGCTAGCTTGTTCATTAATGTGAGTTGAATTAATCTGATTTGTGTCAGTA 29650	
QY		841 GAAAAGATGTGAATACCTTGGAAAGGAGACACATTTTTTTTAAATATATGCTCTGTAATA 900	
Db		29651 GAAAAGATGTGAATACCTTGGAAAGGAGACACATTTTTTTTAAATATATGCTCTGTAATA 29710	
QY		901 CGGATCAGAAGCGAGGTCCCATGAGGACACCTTCGCCCTTAAACATGCTCAACCCCGGC 960	
Db		29711 CGGATCAGAAGCGAGGTCCCATGAGGACACCTTCGCCCTTAAACATGCTCAACCCCGGC 29770	
QY		961 TGCCATAGCTGCGGTGGTCCCTCCAGGTGACTGCTCCGACAAAGGATGCTCTTCAA 1020	

3398	3599	3657	3675	11517	11316
2503	2527	14158	13689	1044	1024
3341	3393	2386	2336	5448	5071
5997	5970	4939	5026	2779	2792
11401	11247	2187	2206	3013	3120
1470	1443	6457	6489	3869	3810
12564	12727	1771	1836	9708	9870
6660	6745	3788	3991	12923	12794
4555	4619	4838	4792	4563	4622
1729	1734	3160	3127	287	<800
406	<800	4088	4313	9519	9493
2213	2193	5832	6201	2229	2229
6922	6981	942	971	2610	2628
4531	4619	4341	4619	5661	5695
1381	1370	9735	9702	576	<800
1386	1370	4886	4792	820	830
1168	1168	15773	15940	3668	3654
15253	15479	398	<800	9025	8972
3613	3599	5122	5509	5092	5071
8174	8090	1336	1324	1026	1024
2074	2086	10995	10821	---	---
714	<800	4329	4313	---	---
6275	6239	16	<800	---	---
3627	3599	1338	1324	---	---
		978	971	---	---
		2203	2206	---	---
		2343	2336	---	---
		489	<800	---	---
		6300	6489	---	---
		---	<800	---	---
		376	---	---	---
		2981	2969	---	---
		Location/Qualifiers			
		1. .161835			
		/organism="Homo sapiens"			
		/mol_type="genomic DNA"			
		/db_xref="taxon:9606"			
		/chromosome="1"			
		/clone="RP5-994D16"			
		/clone_lib="RPC1 human PAC library 5"			
		1. .4			
		/note="Single subclone region"			
		161830. .161835			
FEATURES		source			
misc_feature					
misc_feature					

Db 29771 TGCCATAGCTCGTGGTCCCTCCAAAGTGACTGTCTCCGACAAAAGGGTACGCTCTTCAA 29830
QY 1021 ACCCATACGTTTAAAGCAATTCAGAAACCCCTGGCTGTGCGGACGACTACAGGGCATT 1080
Db 29831 ACCCATACGTTTAAAGCAATTCAGAAACCCCTGGCTGTGCGGACGACTACAGGGCATT 29890
QY 1081 AAGAAAGACGACTCTATGCCCGCGTAATGTCTCAGATCACAGGACGTAATTTGGAG 1140
Db 29891 AAGAAAGACGACTCTATGCCCGCGTAATGTCTCAGATCACAGGACGTAATTTGGAG 29950
QY 1141 CTGGGAGGAGGAGAACCTTTTCTTCCAGGGGGCTAAGCGGTCTTCGAGCCCTTCCA 1200
Db 29951 CTGGGAGGAGGAGAACCTTTTCTTCCAGGGGGCTAAGCGGTCTTCGAGCCCTTCCA 30010
QY 1201 ATCCCGGCTCCGCGCGGTAACTCCCTGCCAGCGTTCCGGCGTCTTTTTCAGCCGA 1260
Db 30011 ATCCCGGCTCCGCGCGGTAACTCCCTGCCAGCGTTCCGGCGTCTTTTTCAGCCGA 30070
QY 1261 GACACAACTGAACTGGGGGCGCCGACGCGCGGTCTCTGTGAAGTCAAGTTC 1320
Db 30071 GACACAACTGAACTGGGGGCGCCGACGCGCGGTCTCTGTGAAGTCAAGTTC 30129
QY 1321 CTCTCTCCCGTCTCTCGGTAATCTATGTTTTCGTGCGGACTCTCTAATTTAGTT 1380
Db 30130 CTCTCTCCCGTCTCTCGGTAATCTATGTTTTCGTGCGGACTCTCTAATTTAGTT 30189
QY 1381 CCGGTCTCTATGCGCGCGCGGAGCAGGACGTTGTAGTCTGACTGTAATAGCGGC 1440
Db 30190 CCGGTCTCTATGCGCGCGCGGAGCAGGACGTTGTAGTCTGACTGTAATAGCGGC 30249
QY 1441 AAGGTCCTAATGAGATGAGGAGCTGATAAAATATTAGCAATAGAGCTAGGGATTGG 1500
Db 30250 AAGGTCCTAATGAGATGAGGAGCTGATAAAATATTAGCAATAGAGCTAGGGATTGG 30309
QY 1501 GTCAGTGGGAGATGACAGTACCACTGGCCAGTGAAACAGCCCTAGGCGGGTCTGCTC 1560
Db 30310 GTCAGTGGGAGATGACAGTACCACTGGCCAGTGAAACAGCCCTAGGCGGGTCTGCTC 30369
QY 1561 GTAGGCTTATCCGCTGTCCGCAATCTCGTGTAGTTCGATCGGTAGCGGAGCGAG 1620
Db 30370 GTAGGCTTATCCGCTGTCCGCAATCTCGTGTAGTTCGATCGGTAGCGGAGCGAG 30429
QY 1621 AGCGGACCCAGAGAGCCCTGAGCAGCCCAACCGCGCGCGGCTAGTACCATCACA 1680
Db 30430 AGCGGACCCAGAGAGCCCTGAGCAGCCCAACCGCGCGCGGCTAGTACCATCACA 30489
QY 1681 CCCCAGGAGGAGCGGAG 1698
Db 30490 CCCCAGGAGGAGCGGAG 30507

RESULT 3

HUMY292E04 HUMY292E04 659 bp mRNA linear PRI 29-AUG-1998
LOCUS Homo sapiens full length insert cDNA clone Y292E04.
DEFINITION AF086102
ACCESSION AF086102.1 GI:3483447
VERSION FLI CDNA.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 659)
AUTHORS Woesner, J., Tan, F., Marra, M., Kucaba, T., Vandell, M., Martin, J.,
Marth, G., Bowles, L., Wylie, T., Bowers, Y., Steptoe, N., Theising, B.,
Geisel, S., Allen, M., Underwood, K., Chappell, J., Person, B.,
Gibbons, M., Harvey, N., Pape, D., Chamberlain, A., Morales, R.,
Schurk, R., Ritter, E., Kohn, S., Swaller, T., Behnmer, K., Hillier, L.,
Wilson, R. and Waterston, R.

TITLE Full Clone Sequencing of the Longest Available Member from Each
Unigene Cluster
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 659)

AUTHORS
TITLE
JOURNAL
COMMENT

Waterston, R.
Direct Submission
Submitted (24-AUG-1998) Department of Genetics, Washington
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:est@watson.wustl.edu

NOTICE: This sequence represents the full insert of this cDNA. No
attempt has been made to verify whether this corresponds to the
full-length of the original mRNA from which it was derived. We
have tried to obtain double-stranded, or double chemistry sequence
across the entire clone, but potentially, there are areas in the
sequence where this level of coverage was not achieved.
Nevertheless, we are confident of the accuracy of this sequence as
all regions of low quality, as defined by PHRAP (P. Green, in
preparation), were visually inspected and edited accordingly. The
consensus quality values for this sequence have been submitted
separately.

The location of this clone is unknown.

FEATURES
source

Location/Qualifiers
1..659
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:290526"
/clone_lib="Soares_multiple_sclerosis_2NbHMSp"
repeat_region
20..119
/rpt_family="L2"

ORIGIN

Query Match 26.98; Score 456.6; DB 9; Length 659;
Best Local Similarity 98.94; Pred. No. 9.2e-126;
Matches 470; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 353 CCCCTCAGGATATATTTCCAAATATAATACCTAAAGTGAAGTCCACCTCTGCAACCCAA 412
Db 177 CACTTGAGGATATATTTCCAAATATAATACCTAAAGTGAAGTCCACCTCTGCAACCCAA 236
QY 413 TGTCAATTCGAGTCTTACTGAACTACTTGAATCTTGAATTCCTGAGATCTCACCTTTCTC 472
Db 237 TGTCAATTCGAGTCTTACTGAACTACTTGAATTCCTGAGATCTCACCTTTCTC 296
QY 473 GCCTGTACCTGTGCGGGAAGTCAAGTCCCTCCACCTTCTCCCTGCTTCCACTCCCAAA 532
Db 297 GCCTGTACCTGTGCGGGAAGTCAAGTCCCTCCACCTTCTCCCTGCTTCCACTCCCAAA 356
QY 533 TACTTCGTGGTTTTCAGCTCTGAGTATTTTACCGTGTGGCTGTGTTTAAATTTCTGCTC 592
Db 357 TACTTCGTGGTTTTCAGCTCTGAGTATTTTACCGTGTGGCTGTGTTTAAATTTCTGCTC 416
QY 593 CATCAGAGGAGAGAACTGACTCGCGAACTATTCCATCCCGAGCCGATAGTAGCGCTTA 652
Db 417 CATCAGAGGAGAGAACTGACTCGCGAACTATTCCATCCCGAGCCGATAGTAGCGCTTA 476
QY 653 AAAAAGAACGGAAGAGGTGGGAGGACTTCAGTAAATCAGGTGGGAGGCTCAATT 712
Db 477 AAAAAGAACGGAAGAGGTGGGT-GGAGGACTTCAGTAAATCAGGTGGGAGGCTCAATT 535
QY 713 TTATCGTTTGTGAACGTTGATAGTATCCCTCTATCAGCTGCTGTGTCAGGAAATAAG 772
Db 536 TTATCGTTTGTGAACGTTGATAGTATCCCTCTATCAGCTGCTGTGTCAGGAAATAAG 595
QY 773 TGAATAAACAACAGAGCTAGCTTGTTCATTAATGTAGTGTGAATTAATCTGA 827
Db 596 TGAATAAACAACAGAGCTAGCTTGTTCATTAATGTAGTGTGAATTAATCTGA 650

```

RESULT 4
BX470216/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

BX470216      162454 bp      DNA      linear      ROD 03-JUL-2003
Mouse DNA sequence from clone RP23-390M8 on chromosome 4, complete
sequence.
ACCESSION     BX470216
VERSION       BX470216.11
KEYWORDS      GI:32451370
SOURCE        HTG.
ORGANISM      Mus musculus (house mouse)
REFERENCE     Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              1 (bases 1 to 162454)
              Pelan, S.
              Direct Submission
              Submitted (03-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              humquery@sanger.ac.uk
              Clone requests: clonerequest@sanger.ac.uk
              On Jul 3, 2003 this sequence version replaced GI:32439594.
              Sequence from the Mouse Genome Sequencing Consortium whole genome
              shotgun may have been used to confirm this sequence. Sequence data
              from the whole genome shotgun alone has only been used where it has
              a phred quality of at least 30.
              ----- Genome Center
              Center: Wellcome Trust Sanger Institute
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: humquery@sanger.ac.uk

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e. phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
EMBL, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPPEP; Information on the WORMPPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-390M8 is from the RPI-23 Mouse BAC Library

constructed by the group of Pieter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6.

```

FEATURES
source
Location/Qualifiers
1. 162454
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-390M8"
/clone_lib="RPCr-23"

```

ORIGIN		Query Match	11.0%; Score 186.8; DB 10; Length 162454;
		Best Local Similarity 70.9%; Pred. No. 8.9e-44;	
		Matches 346; Conservative 0; Mismatches 132; Indels 10; Gaps 7;	
QY	1212	GGCGGGTAATCCCTGCCAGCGTTGGGGGTGCGCTTTTTCAGCGCGAGACACACCT	1271
Db	7620	GC0CGGCGTTACCCAGCGTGGGAGGAGCCCTTCTAGGAGCGCGAGTAGCCCGG	7561
QY	1272	GAAGTGGGGGCGCCAGCGCGCGCTGCCTCGTGGAGTCAAGTTCCTTCTGCGCCGT	1331
Db	7560	CGATCGGGGCGCCAGTTCGGGACGAGCGCGAGCCCATCTGGGACAGTTTCCTCTGCGCGT	7501

QY	1332	CCTCTCGGTA	CTCTATG	TTTTG	TGCGCGG	ACTACTCTAA	TTCTAGT	TTCGGT	CTCTCTAT	13391	
Db	7500	CCTC	-CCGGTACT	CTATG	TGTC	TACGCGCGA	-GGCTCTC	ATTCTAGT	TTCGGTCTCTAT	7443	
QY	1392	GGCGGCGG	CGGAGCAG	GAACGGT	TTGTAGG	TCCAGT	CGAATTA	TACGCCCA	AAAGT-CCAA	1450	
Db	7442	GG----	CGCGGAG	GCCTG	AGAAGCT	TGGGT	CGGCGGA	TTAGT	CGCTAAAGATACCA	7387	
QY	1451	TGAGAAT	TGAGGAG	CTGAT	AAATAT	TAGCCAA	TAGAAGCT	TAGGGA	TTGGGGT	CGGTGGG	1510
Db	7386	TGAGAGCG	CAGGAG	CGCGG	AGCGGT	GAGCAAT	TAGCGCT	TAGAGA	-CCCGGT	CAGGTGGT	7328
QY	1511	CAGATT	TGACAGT	ACCACT	GCCCA	TGAA	CAACCCCT	TAGGCGCGGT	CGCTCGT	AGGCGCTTA	1570
Db	7327	GAGAT	TGACAA	CGAT	TCGGAC	CAAT	TGGGAG	GACGAGG	CGCGGTCTAT	CGCGGGCTTA	7258
QY	1571	TCCCGCGCT	TCGCCGC	CACTTCT	CGCTAGT	TCGAT	CGGTAG	CGGAGCGG	AGCGGAGCGG	ACCCC	1630
Db	7267	TCCCGCGCGT	CCCGCA	CACTTCT	CGCTAGT	TCGAT	CGGTAG	CGGAGCGG	AGCGGAGCGG	ACCCC	7208
QY	1631	AGAGAGCC	CTGAGC	AGCCCA	CCGCGCGC	CGCGGCT	TAGTTA	CAATCACA	CCCCCGG	GAGG	1690
Db	7207	AGAGAGCC	CTGAG-AG	CCCCCA	CCGCGCGC	CGCGGCT	TAGTCA	CCATCACA	-CCCGG	GAGG	7150
QY	1691	AGCCGC	GAG	1698							
Db	7149	AGCCGC	GAG	7142							

RESULT 5					
AC101988/c					
LOCUS	AC101988	194020 bp	DNA	linear	HTG 19-MAR-2003
DEFINITION	Mus musculus clone RP24-363D14,	WORKING DRAFT SEQUENCE, 14			
	unordered pieces.				
ACCESSION	AC101988				
VERSION	AC101988.3	GI:29124242			
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.				
SOURCE	Mus musculus (house mouse)				

REFERENCE	AUTHORS	TITLE	MUS MUSCULUS	ORGANISM
1	(bases 1 to 194020)			
2	(bases 1 to 194020)			
3	(bases 1 to 194020)			
4	(bases 1 to 194020)			
5	(bases 1 to 194020)			
6	(bases 1 to 194020)			
7	(bases 1 to 194020)			
8	(bases 1 to 194020)			
9	(bases 1 to 194020)			
10	(bases 1 to 194020)			
11	(bases 1 to 194020)			
12	(bases 1 to 194020)			
13	(bases 1 to 194020)			
14	(bases 1 to 194020)			
15	(bases 1 to 194020)			
16	(bases 1 to 194020)			
17	(bases 1 to 194020)			
18	(bases 1 to 194020)			
19	(bases 1 to 194020)			
20	(bases 1 to 194020)			
21	(bases 1 to 194020)			
22	(bases 1 to 194020)			
23	(bases 1 to 194020)			
24	(bases 1 to 194020)			
25	(bases 1 to 194020)			
26	(bases 1 to 194020)			
27	(bases 1 to 194020)			
28	(bases 1 to 194020)			
29	(bases 1 to 194020)			
30	(bases 1 to 194020)			
31	(bases 1 to 194020)			
32	(bases 1 to 194020)			
33	(bases 1 to 194020)			
34	(bases 1 to 194020)			
35	(bases 1 to 194020)			
36	(bases 1 to 194020)			
37	(bases 1 to 194020)			
38	(bases 1 to 194020)			
39	(bases 1 to 194020)			
40	(bases 1 to 194020)			
41	(bases 1 to 194020)			
42	(bases 1 to 194020)			
43	(bases 1 to 194020)			
44	(bases 1 to 194020)			
45	(bases 1 to 194020)			
46	(bases 1 to 194020)			
47	(bases 1 to 194020)			
48	(bases 1 to 194020)			
49	(bases 1 to 194020)			
50	(bases 1 to 194020)			
51	(bases 1 to 194020)			
52	(bases 1 to 194020)			
53	(bases 1 to 194020)			
54	(bases 1 to 194020)			
55	(bases 1 to 194020)			
56	(bases 1 to 194020)			
57	(bases 1 to 194020)			
58	(bases 1 to 194020)			
59	(bases 1 to 194020)			
60	(bases 1 to 194020)			
61	(bases 1 to 194020)			
62	(bases 1 to 194020)			
63	(bases 1 to 194020)			
64	(bases 1 to 194020)			
65	(bases 1 to 194020)			
66	(bases 1 to 194020)			
67	(bases 1 to 194020)			
68	(bases 1 to 194020)			
69	(bases 1 to 194020)			
70	(bases 1 to 194020)			
71	(bases 1 to 194020)			
72	(bases 1 to 194020)			
73	(bases 1 to 194020)			
74	(bases 1 to 194020)			
75	(bases 1 to 194020)			
76	(bases 1 to 194020)			
77	(bases 1 to 194020)			
78	(bases 1 to 194020)			
79	(bases 1 to 194020)			
80	(bases 1 to 194020)			
81	(bases 1 to 194020)			
82	(bases 1 to 194020)			
83	(bases 1 to 194020)			
84	(bases 1 to 194020)			
85	(bases 1 to 194020)			
86	(bases 1 to 194020)			

Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouthgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazarro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Deaellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Gatagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamarez, R., Landers, T., Lehoczyk, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneses, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollars, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triglio, J., Vassiliev, H., Vael, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 194020)

QY 1511 CAGATTACAGTACCACTGGCCAGTGAACAACGCTTAGGGCGGTCTCGTAGGGCTTA 1570
 DB 120081 GAGATGACAAACGATCGGACCAATCGGGGAGGACGAGGGCGGTCTATCGGGGGCTTA 120022
 QY 1571 TCCCGCTGTCGCCCGCATTCCTGCTAGTTCGATCGGTAGCGGAGCGGAGCGGAGCGGACCCC 1630
 DB 120021 TCCCGCGCTGCCCGCATTCCTGCTAGTTCGATCGGTAGCGGAGCGGAGCGGAGCGGACCCC 119862
 QY 1631 AGAGAGCCCTGAGACGACCCCGCCCGCGCGGCTAGTTCACATCACACCCCGGGAGG 1690
 DB 119961 AGAGAGCCCTGAG-AGCCCCACCGCGCGCGGCTAGTTCACATCAC-CCCGGGAGG 119904
 QY 1691 AGCGCGAG 1698
 DB 119903 AGCGCGAG 119896

RESULT 6
 AC098918/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-187N5, *** SEQUENCING IN PROGRESS

 ACCESSION AC098918
 VERSION AC098918.7 GI:30581384
 KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS ENRICHED.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 225308)
 REFERENCE
 AUTHORS Muzny,D.,Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,P., Biswal,D., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhaty,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K., Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Cantar,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,P., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Loulsged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mathew,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K., Nwaokeme,O., Okwunonu,G., Olarnpusagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,P., Ponderdexter,A., Popovic,D., Primus,E., Fu,L.-L., Puzo,M., Quiroz,J., Rachin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P., Rives,C., Rodkey,T., Rojasa,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbey,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,

Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P., Williams,G., Willson,R., Wlezyk,R., Woodson,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhauser,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission
 Unpublished
 2 (bases 1 to 225308)
 Worley,K.C.
 Direct Submission
 Submitted (06-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 225308)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On May 13, 2003 this sequence version replaced gi:24942594.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GILR
 Center clone name: CH230-187N5
 ----- Summary Statistics
 Assembly program: Atlas 3.0;
 Consensus quality: 210466 bases at least Q40
 Consensus quality: 213831 bases at least Q30
 Consensus quality: 216522 bases at least Q20
 Estimated insert size: 224754; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: this is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 225308: contig of 225308 bp in length.

----- Location/Qualifiers
 1 .225308
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-187N5"
 misc_feature 34231..36103
 /note="wgs_contig"

FEATURES
 source
 misc_feature
 ORIGIN

Query Match	11.0%;	Score 186.2;	DB 2;	Length 225308;
Best Local Similarity	79.7%;	Pred. No. 1.5e-43;		
Matches	307;	Conservative	0;	Mismatches 68; Indels 10; Gaps 7;
Qy	1315	ACGTTCCCTTCGCGGTCTCTCTATGCGCGCCCGGAGGCGAGAACGGTTTGTAGGTCGACTGAATTA	1374	
Db	207314	ACGTTCCCTTCGCGGTCTCTCTATGCGGTACTCTATGGTTTTCGTGGCCGACTACTCTAATT		207257
Qy	1375	CTAGTTCCGGTCTCTATGCGCGCCCGGAGGCGAGAACGGTTTGTAGGTCGACTGAATTA	1434	
Db	207256	CTAGTTCCGGTCTCTATGG-----CGGCGGAGGCGGAGAGGCTGTGGTCGGCGGAATTA		207201
Qy	1435	GCGGCCAAGGT-CCAAATGAGATGAGAGACTGATAAATATTAGCCAAATAGAAGCTAGG	1493	
Db	207200	GTCGCTAAAGGTACCAATGAGAGCACAGGAGCCGCGAGCGTCTAGCCAAATAGGTGCTAGA		207141
Qy	1494	GATTGGGGTCAGGTGGGAGATTGACATGACCACTGGCCAGTGAACACGCCCTTAGGCGG	1553	
Db	207140	GA-CCCGGTCAGGTGTGATGTGGAACACGATGTAGGCCAATGGGAGAGGACGAGGCGG		207082
Qy	1554	GTCGCTCGTAGGGCTATTCCCGCCTGTCCCGCCCAATCTCGCTAGTTTCGATCGGTACGCGG	1613	
Db	207081	GTTACCGCGGGGCTATTCCCGCCGTCCCGCCCAATCTCGCTAGTTTCGATCGGTACGCGG		207022
Qy	1614	AGCGGAGAGCGGACCCGAGAGAGCCCTGAGCAGCCCCACCGCCGCGCGGCTAGTTAC	1673	
Db	207021	AGCGGAGAGCGGACCCGAGAGAGCCCTGAG-AGCCCCACCGCGCGGCGCTAGTCAC		206963
Qy	1674	CATCACACCCGGGAGGACCCGAG	1698	
Db	206962	CATCAC-CCCGGAGGAGCGCGAG	206939	
RESULT 7				
AX897942				
LOCUS	AX897942	178 bp	DNA	linear
DEFINITION	Sequence 13805 from Patent EP1033401.			PAT 18-DEC-2003
ACCESSION	AX897942			
VERSION	AX897942.1	GI:40052855		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1			
AUTHORS	Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.			
TITLE	Expressed sequence tags and encoded human proteins			
JOURNAL	Patent: EP 1033401-A 13805 06-SEP-2000;			
FEATURES				
source	Location/Qualifiers			
	1..178			
	/organism="Homo sapiens"			
	/mol_type="unassigned DNA"			
	/db_xref="taxon:9606"			
ORIGIN				
Query Match	6.5%;	Score 111;	DB 6;	Length 178;
Best Local Similarity	99.1%;	Pred. No. 1.1e-21;		
Matches	111;	Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
Qy	1587	ATTCTCGCTAGTTCGATCGGTAGCGGAGCGGAGCGGACCCGAGAGCCCTGAGCAG	1646	
Db	1	ATTCTCGCTAGTTCGATCGGTAGCGGAGCGGAGCGGACCCGAGAGCCCTGAGCAG		60
Qy	1647	CCCCACCGCGCGCGGCGCTAGTTACATCACACCCCGGAGAGCCGCGAG	1698	
Db	61	CCCCACCGCGCGCGGCGCTAGTTACATCACACCCCGGAGAGCCGCGAG		112
RESULT 8				
BD033475				
LOCUS	BD033475	178 bp	DNA	linear
DEFINITION	Sequence tag and encoded human protein.			PAT 27-AUG-2002

Matches 108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1590 CTCGCTAGTTCGATCGGTAGCGGAGCGGAGACCGCCAGAGAGCCCTGAGCAGCCC 1649

Db 1 CTCGCTAGTTCGATCGGTAGCGGAGCGGAGACCGCCAGAGAGCCCTGAGCAGCCC 60

QY 1550 CACCGCGCGCGCGCGCTAGTTCACCATCACACCGCGGAGGAGCCGCGAG 1698

Db 61 CACCGCGCGCGCGCGCGCTAGTTCACCATCACACCGCGGAGGAGCCGCGAG 109

RESULT 10

BD203713

LOCUS

DEFINITION Human nucleic acid sequence originating in ovarian myoma tissue.

ACCESSION

BD203713

VERSION

BD203713.1 GI:33013483

KEYWORDS

JP 2002511252-A/54

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 1548)

AUTHORS

Specht,T., Hinzmann,B., Schmitt,A., Pilarsky,C., Dahl,E. and Rosenthal,A.

TITLE

Human nucleic acid sequence originating in ovarian myoma tissue

JOURNAL

Patent: JP 2002511252-A 54 16-APR-2002;

COMMENT

METAGEN GESELLSCHAFT FUER GENOME FORSCHUNG MBH

OS: Homo sapiens (human)

PN JP 2002511252-A/54

PD 16-APR-2002

PF 07-APR-1999 JP 2000543588

PR 09-APR-1998 DE 198 17 557.4

PI THOMAS SPECHT, BERND HINZMANN, ARMIN SCHMITT, CHRISTIAN PILARSKY,

PI EDGAR DAHL,

PI ANDRE ROSENTHAL

PC C12N15/09,A61K38/00,A61K48/00,A61P15/00,A61P35/00,C07K14/82,

PC C07K16/32,

PC C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12P21/08,C12Q1/68,G01N33/

PC 574,

PC C12N15/00,A61K37/02,C12N5/00

CC Human nucleic acid sequence originating in ovarian myoma CC

tissue

FH Key Location/Qualifiers

FT source 1..1548

FT /organism="Homo sapiens (human)"

Location/Qualifiers

1..1548

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

ORIGIN

Query Match 6.3%; Score 107.4; DB 6; Length 1548;

Best Local Similarity 99.1%; Pred. No. 2.4e-20;

Matches 108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1590 CTCGCTAGTTCGATCGGTAGCGGAGCGGAGACCGCCAGAGAGCCCTGAGCAGCCC 1649

Db 1 CTCGCTAGTTCGATCGGTAGCGGAGCGGAGACCGCCAGAGAGCCCTGAGCAGCCC 60

QY 1650 CACCGCGCGCGCGCGCTAGTTCACCATCACACCGCGGAGGAGCGGCGAG 1698

Db 61 CACCGCGCGCGCGCGCGCTAGTTCACCATCACACCGCGGAGGAGCGGCGAG 109

RESULT 11

AC016114

LOCUS

DEFINITION Human sapiens clone RP11-27C17, WORKING DRAFT SEQUENCE, 8 unordered

ACCESSION

AC016114

VERSION

AC016114.3 GI:7382202

HTG; HTGS PHASE1; HTGS_DRAFT.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 144902)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens, clone RP11-27C17

Unpublished

2 (bases 1 to 144902)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgaiter,B., Brown,A., Castelle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lechoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 1, 2000 this sequence version replaced gi:16587892.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4680

Center clone name: 27_C17

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 138527 bases at least Q40

Consensus quality: 141533 bases at least Q30

Consensus quality: 142759 bases at least Q20

Insert size: 145000; agarose-fp

Insert size: 144202; sum-of-contigs

Quality coverage: 4.6 in Q20 bases; agarose-fp

Quality coverage: 4.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1664: contig of 1964 bp in length

* 1965 2064: gap of 100 bp

* 2065 5122: contig of 3058 bp in length

* 5123 5222: gap of 100 bp

* 5223 14250: contig of 9028 bp in length

* 14251 14350: gap of 100 bp

* 14351 20405: contig of 6055 bp in length

* 20406 35724: contig of 15219 bp in length

* 35725 35824: gap of 100 bp

* 35825 65894: contig of 29770 bp in length

* 65895 103395: contig of 37701 bp in length

* 103396 103495: gap of 100 bp

Contact: SeqRef@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : C-2132D14 PCR1 Downstream
BAC (overlapping the SP6 end) : R-783L4 (AC=AL356021)

----- Summary Statistics -----
Assembly program: Phrap; version 2.0
Quality coverage: 6.46x in Q20 bases; sum-of-contigs

----- Overall quality chart -----
Range : bases
0 :
1 - 9 :
10 - 19 : 9
20 - 29 : 32
30 - 39 : 224
40 - 49 : 5128
50 - 59 : 8470
60 - 69 : 10629
70 - 79 : 22286
80 - 89 : 51219
90 - 99 : 64147

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES
source
Location/Qualifiers
1..162144
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/clone="C-2014B16"
/clone.lib="CalTech-D"
107873..108073
/note="matching EMBL:L16415
RDB:RH11054
dbSTS:STS14862
Identified using the e-PCR software (G. Schuler)"

ST5

ORIGIN

Query Match 6.0%; Score 101.4; DB 9; Length 162144;
Best Local Similarity 99.0%; Pred. No. 5.le-18;
Matches 102; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1596 AGTTGCATCGGTAGCGGAGCGGAGCGGAGCCGCCAGAGAGCCCTGAGCAGCCCCACC GC 1655
Db 5451 AGTTGCATCGGTAGCGGAGCGGAGCGGAGCCGCCAGAGAGCCCTGAGCAGCCCCACC GC 5510
QY 1656 CGCCCGCGCGCTAGTTCATCATCACACCCCGGAGGAGCGCGAG 1698
Db 5511 CGCCCGCGCGCTAGTTCATCATCACACCCCGGAGGAGCGCGAG 5553

RESULT 13
CN306CB8
LOCUS
DEFINITION Human chromosome 14 DNA sequence BAC R-783L4 of library RPCI-11
from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION AL356021
VERSION AL356021.4 GI:14572588
KEYWORDS HTG; HTGS; ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 184656)
Heilig.R., Petit.J.L., Vico.V., Dasilva.C., Robert,C., Wincker,P.,
Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
Cyapay,G., Saurin,W. and Weissenbach,J.
Sequencing of the human chromosome 14
Unpublished
TITLE JOURNAL Unpublished
AUTHORS 2 (bases 1 to 184656)
JOURNAL Unpublished
COMMENT On Apr 30, 2001 this sequence version replaced gi:13016612.

* 103496 144902: contig of 41407 bp in length.
Location/Qualifiers
1..144902
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-27C11"
/clone.lib="RPCI-11 Human Male BAC"
1..1964
/note="assembly_fragment"
2065..5122
/note="assembly_fragment"
5223..14250
/note="assembly_fragment"
14351..20405
/note="assembly_fragment"
clone.end.SP6
vector_side:left
20506..35724
/note="assembly_fragment"
35825..65594
/note="assembly_fragment"
clone.end.T7
vector_side:right
65695..103395
/note="assembly_fragment"
103496..144902
/note="assembly_fragment"

ORIGIN

Query Match 6.0%; Score 101.4; DB 2; Length 144902;
Best Local Similarity 99.0%; Pred. No. 4.9e-18;
Matches 102; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1596 AGTTGCATCGGTAGCGGAGCGGAGCGGAGCCGCCAGAGAGCCCTGAGCAGCCCCACC GC 1655
Db 137738 AGTTGCATCGGTAGCGGAGCGGAGCGGAGCCGCCAGAGAGCCCTGAGCAGCCCCACC GC 137797
QY 1656 CGCCCGCGCGCTAGTTCATCATCACACCCCGGAGGAGCGCGAG 1698
Db 137798 CGCCCGCGCGCTAGTTCATCATCACACCCCGGAGGAGCGCGAG 137840

RESULT 12
CN306CB8
LOCUS
DEFINITION Human chromosome 14 DNA sequence BAC C-2014B16 of library CalTech-D
from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION AL391261
VERSION AL391261.3 GI:13897302
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 162144)
Heilig.R., Petit.J.L., Vico.V., Dasilva.C., Robert,C., Wincker,P.,
Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
Cyapay,G., Saurin,W. and Weissenbach,J.
Sequencing of the human chromosome 14
Unpublished
TITLE JOURNAL Unpublished
AUTHORS 2 (bases 1 to 162144)
Direct Submission
Submitted (02-MAY-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On Apr 30, 2001 this sequence version replaced gi:13016612.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/

AL152733
AL152733.9 GI:15028630
HTG.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Cho
Mammalia; Eutheria; Pri
1 (bases 1 to 189367)
Stuce, C.
Direct Submission
Submitted (15-OCT-2001)

On Jul 29, 2001 this sequence version replaced gi:14329913.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em1; EMBL; SW1; SWISSPROT; Tr1; TREMBL; Wp1; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep. This sequence was generated from pat of bacterial clone contigs of human

Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>
RP11-377P23 is from the library RPC1-11.2 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>

```
Location/Qualifiers
1..189367
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-397F23"
/clone_1kb="RP11-397F23"
/clone_1kb="RP11-397F23"
```

short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."

74553. .75378

/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."

Query Match	5.7%	Score 96.6;	DB 9;	Length 189367;
Best Local Similarity	96.1%	Pred. No. 1.5e-16;		
Marches	99;	Conservative	0;	Mismatches 4;
			Indels	0;
			Gaps	0;

Best Local Similarity 96.1%; Pred. No. 1.5e-16;
 Mismatches 0; Conservative 4; Indels 0; Gaps 0;

Qy	1596	AGTTTCGATCGTTACGGGGACGGAGCGAGCCCAAGAGAGCCTCTGAGCAGGCCACCGC
Db	111342	AGTTTCGATCGTTACGGGACGGAGCGGACCCCAAGAGAGCCTCTGAGCAGGCCCCACAC
Ov	1656	CGCCGCGCGGCTAGTTATTCATCAACCCCGGAGAGCGCGCAG

	Query Match	5.6%	Score 95;	DB 9;	Length 99886;
	Best Local Similarity	95.1%;	Pred. No. 3.8e-16;		
	Matches 98;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;
Qy	1596	AGTTCGATCGGTTAGCGGAGCGGAGACCGACCCACAGAGCCCTGACGACGCCACGCCACCGC	1555		
Db	97059	AGTTCAATCGGTAGTCGGAGCGGAGAGCTGACCCACAGAGACCTGTGGCAGCCGCCACCTC			
Qy	1556	CGCGCGCGGCGCTAGTTTACCATCATCACCCGCGGAGGACCGCAG	1698		
Db	96999	CGCGCGCGGCGCTAGTTTACCATCATCACCCGCGGAGAGACCGCAG	96957		

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 12:38:54 ; Search time 755 Seconds
(without alignments)
9554.226 Million cell updates/sec

Title: US-09-869-508-1_COPY_453_2150

Perfect score: 1698

Sequence: 1 agaacatatgtcggccaa.....caccgccggagagcgccgac 1698

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002s:*
- 7: Geneseqn2003as:*
- 8: Geneseqn2003bs:*
- 9: Geneseqn2003cs:*
- 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	469.6	27.7	2102	5	AAS83003 DNA encod
2	111	6.5	178	3	AAC09730 Human sec
3	107.4	6.3	1548	2	Aaz77507 Human ova
4	95	5.6	1978	7	AAL51561 Human nuc
5	89.4	5.3	395	6	AAS61724 Lung smal
6	89.4	5.3	395	9	ADD66913 Human lun
7	89.4	5.3	395	9	ADBE8167 Human lun
8	89	5.2	1481	4	AAC81319 Human Y-b
9	89	5.2	1481	6	AB554998 Human YB-
10	89	5.2	1481	6	ABN55602 Gene #210
11	89	5.2	1481	9	ABZ99616 Human Y-b
12	88.6	5.2	179	4	AAL02684 Human rep
13	87.4	5.1	1520	9	ADB47382 Human CDN
14	85.2	5.0	410	4	AAD23430 Human lun
15	85.2	5.0	410	9	ADD66708 Human lun
16	85.2	5.0	410	9	ADE87962 Human lun
17	83	4.9	3073	2	AAL04635 Clone ass
18	83	4.9	3073	2	AAT34371 Plasmid p
19	83	4.9	3073	2	Aaz32246 Human gli
20	83	4.9	3073	3	Aaz32246 Human gli
21	81.4	4.8	1474	6	ABK84103 Human CDN
22	78.8	4.6	381	3	AAC65939 Human lun
23	78.8	4.6	381	6	ABL49158 Human lun

24	78.8	4.6	381	6	ABQ92344	Abq92344 Human lun
25	78.8	4.6	381	8	ADA28305	Ada28305 Human lun
26	78.8	4.6	381	9	AD53510	Ad53510 Human lun
27	77.4	4.6	545	6	AAS61751	Aas61751 Lung smal
28	77.4	4.6	545	9	ADD66940	Add66940 Human lun
29	77.4	4.6	545	9	AD588194	Ad588194 Human lun
30	73.4	4.3	1516	7	AAD55839	Aad55839 Human nuc
31	61.2	3.6	2179	4	AAH15631	Aah15631 Human CDN
32	47.4	2.8	60	6	ABN49420	Abn49420 Human spl
33	46.4	2.7	2000	7	ADA71938	Ada71938 Rice gene
34	45.2	2.7	3900	4	AAS26731	Aas26731 Human gen
35	45.2	2.7	3900	7	ABX74080	Abx74080 Human nov
36	45.2	2.7	3901	4	AAS26729	Aas26729 Human nov
37	45.2	2.7	3901	7	ABX74078	Abx74078 Human nov
38	44.8	2.6	3152	4	AAS28912	Aas28912 Human imm
39	44.8	2.6	3152	9	ADB31753	Adb31753 Human nov
40	41.8	2.5	1469	6	AAN92735	Aan92735 Sheep PrP
41	41.6	2.4	1469	6	ABI99233	Abi99233 Mouse isc
42	41.4	2.4	10732	3	AAAI0594	Aaai0594 Gene enco
43	41.2	2.4	2452	6	AAI71445	Aai71445 TNFR/NGER
44	41	2.4	602	4	RAI37306	Rai37306 Human mus
45	41	2.4	602	7	ABX60294	Abx60294 cDNA enco

ALIGNMENTS

RESULT 1

AAS83003

ID AAS83003 standard; cDNA; 2102 BP.

AC AAS83003;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #18807.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; es.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

P-PSDB; ABG18816.

New isolated polynucleotide and encoded polypeptides, useful in

diagnostics, forensics, gene mapping, identification of mutations

responsible for genetic disorders or other traits and to assess

biodiversity.

Claim 1; SEQ ID NO 18807; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (ii) and its binding partners are useful in medical imaging
CC of sites expressing (ii). (i) and (ii) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences

XX SQ Sequence 2102 BP; 580 A; 540 C; 547 G; 434 T; 0 U; 1 Other;

Query Match 27.7%; Score 469.6; DB 5; Length 2102;

Best Local Similarity 97.8%; Pred. No. 3.6e-138;

Matches 497; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

QY 837 AGTAGAAAAGATGTGAATACCTTGGAAAGGAGACACATTTTTTAAATATATGCTGT 896

Db 1 AGTAGAAAAGATGTGAATACCTTGGAAAGGAGACACATTTTTTAAATATATGCTGT 60

QY 897 AAAACGGATCAGAACGAGTCCCATGAGACACACCTCGCCCTAAACATCTGAACCC 956

Db 61 AAAACGGATCAGAACGAGTCCCATGAGACACACCTCGCCCTAAACATCTGAACCC 120

QY 957 GGCTGCCATAGCTCGGTGCTCCCTCAAGGTGACTGCTCGACAAAAGGTACGCTCT 1016

Db 121 GGCTGCCATAGCTCGGTGCTCCCTCAAGGTGACTGCTCGACAAAAGGTACGCTCT 180

QY 1017 TCAACGCCA-TAGCTTTAAGCAATTCAGAACCCCTCGCTGTGCGGCGACTTACACGGC 1075

Db 181 TCAACGCCA-TAGCTTTAAGCAATTCAGAACCCCTCGCTGTGCGGCGACTTACACGGC 240

QY 1076 CATTAAAGAAAAGACGACTCTATGCCCGCGTAATGTTCTAGATCAGAGGACGTAAT 1135

Db 241 CATTAAAGAAAAGACGACTCTATGCCCGCGTAATGTTCTAGATCAGAGGACGTAAT 300

QY 1136 TGGAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1195

Db 301 TGGAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360

QY 1196 TTCCATCCCGGTTCGGGCGGGTAATCCCTGCCAGCGTTGCGGGGTGCTTTTTCATCA 1255

Db 361 TTCCATCCCGGTTCGGGCGGGTAATCCCTGCCAGCGTTGCGGGGTGCTTTTTCATCA 420

QY 1256 GCCGAGACACACCTGAACTGGGGGCGCCGCGAGCCCGCGGCTGCTCGTGGAAATCA 1315

Db 421 GCCGAGACTCAACCTGAACTG-GGGGCGCGCGAGCCCGAGCGCTGCTAGTGAATCA 479

QY 1316 CGTTCCTTCTGCGGCTCTCGGGTAC 1343

Db 480 CGTTCCTTCTGCGGCTCTCGGGTAC 507

RESULT 2

AAC09730

ID AAC09730 standard; cDNA; 178 BP.

XX AAC09730;

AC AAC09730;

XX 06-OCT-2000 (first entry)

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 13805.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX gene therapy; chromosome mapping; ss.

KW Homo sapiens.

OS Homo sapiens.

XX EP1033401-A2.

XX

FN

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-00200610.

XX 26-FEB-1999; 99US-0122487P.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

XX diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX Claim 1; SEQ ID NO 13805; 71pp + Sequence Listing; English.

XX The present sequence is one of a large number of 5' ESTs derived from

XX mRNAs encoding secreted proteins. No ORF has yet been conclusively

XX identified within the present sequence. The 5' ESTs were prepared from

XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST

XX sequences usually correspond mainly to the 3' untranslated region (UTR)

XX of the mRNA because they are often obtained from oligo-dT primed cDNA

XX libraries. Such ESTs are not well suited for isolating cDNA sequences

XX derived from the 5' ends of mRNAs and even in those cases where longer

XX cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'

XX ESTs are derived from mRNAs with intact 5' ends and can therefore be used

XX to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in

XX diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX They are used to obtain upstream regulatory sequences and to design

XX expression and secretion vectors

XX SQ Sequence 178 BP; 38 A; 70 C; 52 G; 17 T; 0 U; 1 Other;.

Query Match 6.5%; Score 111; DB 3; Length 178;

Best Local Similarity 99.1%; Pred. No. 1.7e-24;

Matches 111; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1587 ATTCTCGCTAGTTCGATCGGTAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 1646

Db 1 ATTCTCGCTAGTTCGATCGGTAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 60

QY 1647 CCCACCGCGCGCGCGCGCTAGTTACCATCACACCCCGGAGGAGCGCGAG 1698

Db 61 CCCACCGCGCGCGCGCGCTAGTTACCATCACACCCCGGAGGAGCGCGAG 112

RESULT 3

AAC097507

ID AAC097507 standard; cDNA; 1548 BP.

XX AAC097507;

XX 10-APR-2000 (first entry)

DT 10-APR-2000 (first entry)

DE Human ovarian tumor cDNA library derived EST fragment 58.

XX Expressed sequence tag; EST; human; ovarian tumor; anticancer;

XX gene therapy; treatment; ss.

XX Homo sapiens.

XX DE19817557-A1.

XX 21-OCT-1999.

XX 09-APR-1998; 98DE-01017557.

XX 09-APR-1998; 98DE-01017557.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX

PA


```

PR 21-JUN-2000; 2000US-0213361P.
PR 01-SEP-2000; 2000US-0229763P.
PR 05-SEP-2000; 2000US-0230629P.
PR 14-SEP-2000; 2000US-0232565P.
PR 19-DEC-2000; 2000US-0257037P.
PR 08-JAN-2001; 2001US-0260796P.
XX (CORI-) CORIXA CORP.
PA
XX
PI Lodes MJ, Wang T, Mohamath R, Indirias CY;
XX
XX WPI; 2002-010896/01.
XX
XX Lung tumor polynucleotide and polypeptides useful in therapy and
PT diagnosis of cancer especially lung cancer.
XX
XX Claim 1; Page 226; 295pp; English.
XX
XX The invention relates to novel isolated lung small cell cancer antigen
CC polynucleotides (I) and polypeptides (II) used in a method of detecting
CC cancer in a patient. The method is optionally performed by utilising
CC oligonucleotides (III), where the biological sample from the patient is
CC contacted with (III), detecting the amount of polynucleotide hybridised
CC to (III) in the sample and comparing the amount of polynucleotide to a
CC predetermined cut-off value and thereby determining cancer in a patient.
CC (I), (II) or antigen-presenting cells expressing (II) is useful for
CC stimulating and/or expanding T cells specific for a tumour protein. The
CC method comprises contacting T cells with one of the components under
CC conditions to permit the stimulation and/or expansion of the cells. A
CC composition comprising (I) is useful for stimulating an immune response
CC in a patient and for inhibiting the development of a cancer especially
CC lung cancer in a patient. An isolated T cell population is useful for
CC removing tumour cells from the biological sample and for inhibiting the
CC development of cancer in a patient. AAS61460-AAS61874 represent novel
CC human lung small cell cancer antigen coding sequences of the invention
XX
XX Sequence 395 BP; 86 A; 138 C; 126 G; 44 T; 0 U; 1 Other;
SQ
Query Match 5.3%; Score 89.4; DB 6; Length 395;
Best Local Similarity 98.9%; Pred. No. 2.2e-17;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1608 AGCGGGAGCGGAGCGGACCCAGAGAGCGCTGAGAGAGCCGCCACCGCGCGCGGCT 1667
DB 13 AGGGGAGCGGAGCGGACCCAGAGAGCGCTGAGAGAGCCGCCACCGCGCGCGGCT 72
QY 1668 AGTTACCATCACACCCCGGAGGAGCGCGAG 1698
DB 73 AGTTACCATCACACCCCGGAGGAGCGCGAG 103
RESULT 6
ADD66913
ID ADD66913 standard; cDNA; 395 BP.
XX
XX ADD66913;
XX
XX 15-JAN-2004 (first entry)
XX
XX Human lung tumour-specific related cDNA, SEQ ID No 605.
XX
XX expression control; cancer; T cell; tumour; immune; cytostatic; vaccine;
KW human; lung tumour-specific; ss.
XX
XX Homo sapiens.
XX
XX WO200292001-A2.
XX
XX 21-NOV-2002.
XX
XX 10-MAY-2002; 2002WO-US014975.
XX
XX 11-MAY-2001; 2001US-00854133.
XX

```

```

XX (CORI-) CORIXA CORP.
XX
XX Lodes MJ, Wang T, Fan L, Algate PA, Mcneill PD;
XX
XX WPI; 2003-120592/11.
XX
XX New polynucleotide and polypeptide, useful for preparing a composition
PT for diagnosing, treating or preventing cancer.
XX
XX Claim 11; SEQ ID NO 605; 494pp; English.
XX
XX The invention relates to a novel isolated polynucleotide comprising one
CC of 32 47-6080 base pair sequences, given in the specification, or their
CC complements or degenerate variants, at least 20 contiguous residues of a
CC sequence in, or having at least 75 or 90 % identity with the isolated
CC polynucleotide, or that hybridise with the polynucleotide. The invention
CC further comprises: an isolated polypeptide; an expression vector
CC comprising the polynucleotide operably linked to an expression control
CC sequence; a host cell transformed or transfected with the expression
CC vector; an isolated antibody or its antigen-binding fragment that
CC specifically binds to the polypeptide; a method for detecting the
CC presence of a cancer in a patient; a fusion protein comprising the
CC polypeptide; an oligonucleotide that hybridises to the isolated
CC polynucleotide under moderately stringent conditions; a method for
CC stimulating and/or expanding T cells specific for a tumour protein; an
CC isolated T cell population; a composition comprising a first component
CC consisting of carriers and immunostimulants and a second component; a
CC method for stimulating an immune response in a patient; a method for
CC treating cancer in a patient; a method for determining cancer in a
CC patient; a diagnostic kit comprising at least one oligonucleotide or
CC antibody and a detection reagent comprising a reporter group; and a
CC method for inhibiting the development of cancer in a patient. The
CC compositions of the invention have cytostatic activity and can be used to
CC create a vaccine. The isolated polynucleotide is useful for preparing a
CC composition for diagnosing, treating or preventing cancer. This
CC polynucleotide sequence represents a human lung tumour-specific cDNA
CC sequence relating to the invention.
XX
XX Sequence 395 BP; 86 A; 138 C; 126 G; 44 T; 0 U; 1 Other;
SQ
Query Match 5.3%; Score 89.4; DB 9; Length 395;
Best Local Similarity 98.9%; Pred. No. 2.2e-17;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1608 AGCGGGAGCGGAGCGGACCCAGAGAGCGCTGAGAGAGCCGCCACCGCGCGCGGCT 1667
DB 13 AGGGGAGCGGAGCGGACCCAGAGAGCGCTGAGAGAGCCGCCACCGCGCGCGGCT 72
QY 1668 AGTTACCATCACACCCCGGAGGAGCGCGAG 1698
DB 73 AGTTACCATCACACCCCGGAGGAGCGCGAG 103
RESULT 7
ADE88167
ID ADE88167 standard; cDNA; 395 BP.
XX
XX ADE88167;
XX
XX 29-JAN-2004 (first entry)
XX
XX Human lung tumour antigen cDNA #500.
XX
XX Human; lung tumour antigen; gene; ss; cancer; lung cancer; CD4+; CD8+;
KW T cell; immune response; immunostimulant; cytostatic.
XX
XX Homo sapiens.
XX
XX US2003118599-A1.
XX
XX 26-JUN-2003.
XX

```


QY 1610 CGGGAGCGGAGCGGACCCCGAGAGCCCTGAGCAGCCCGCCAGCCCGCGCGCGCTAG 1669
DB 2 CGGGAGCGGAGCGGACCCCGAGAGCCCTGAGCAGCCCGCCAGCCCGCGCGCGCTAG 61

QY 1670 TTACCATCACACCCCGGAGGAGCCGAG 1698
DB 62 TTACCATCACACCCCGGAGGAGCCGAG 90

RESULT 9
AB954998
ID ABS54998 standard; DNA; 1481 BP.
XX
AC ABS54998;
DT 10-DEC-2002 (first entry)
XX
DE Human YB-1 Y Box Binding Protein gene.
XX
KW Telomerase reverse transcriptase; TERT; replication-conditional virus;
KW adenovirus replication gene; cancer cell; lung; pancreatic cancer;
KW medulloblastoma; cervical carcinoma; fibrosarcoma; osteosarcoma;
KW cytolytic; replication defective adenovirus vector; congenital defect;
KW proinflammatory; antiinflammatory; heterologous effector gene;
KW cancer therapy; cytostatic; gene therapy; human; YB-1;
KW Y Box Binding Protein; Gene; ds.
XX
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT CDS 127..1080
FT /*tag= a
FT /product= "Human YB-1 Y Box Binding Protein"

WO200253760-A2.
11-JUL-2002.
XX
XX 17-DEC-2001; 2001WO-US048785.
XX
XX 18-DEC-2000; 2000US-0256418P.
XX
XX (GERO-) GERON CORP.
XX
XX Irving JW, Lebkowski JS;
PI WPI; 2002-723123/78.
DR P-FSB; ABG70989.
XX

Novel replication-conditional virus useful for cytotoxicity of target cells
e.g. cancer cells and preparing a medicament for treating cancer,
comprises heterologous replication element in an adenovirus-based
construct.
XX
XX Disclosure; Page 30; 32pp; English.

The present invention relates to a new replication-conditional virus with
a genome comprising adenovirus replication genes and one or more
heterologous gene(s) that functionally replaces one or more adenovirus
gene(s) required for replication or assembly of the virus. The invention
is useful for killing a cancer cell (such as lung, pancreatic cancer,
medulloblastoma, cervical carcinoma, fibrosarcoma or osteosarcoma),
killing a cell expressing TERT (telomerase reverse transcriptase), and in
preparing a medicament for treating cancer and a condition associated
with increased expression of TERT in affected cells, in a subject. The
invention is also useful for cytotoxicity of specific target cells. The
invention is further useful for producing replication defective
adenovirus vector which is useful for transient expression of a
heterologous therapeutic gene to correct a congenital defect, introducing
proinflammatory or antiinflammatory activity, enhancing telomerase
function, and delivering heterologous effector genes that induce killing
of the transduced cells. The invention is more safe for use in cancer
therapy. The present nucleic acid sequence encodes the human YB-1 Y Box

CC Binding Protein as described in the invention
XX
SQ Sequence 1481 BP; 437 A; 381 C; 386 G; 277 T; 0 U; 0 Other;

Query Match 5.2%; Score 89; DB 6; Length 1481;
Best Local Similarity 100.0%; Pred. No. 7.1e-17;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1610 CGGGAGCGGAGCGGACCCCGAGAGCCCTGAGCAGCCCGCCAGCCCGCGCGCTAG 1669
DB 2 CGGGAGCGGAGCGGACCCCGAGAGCCCTGAGCAGCCCGCCAGCCCGCGCGCTAG 61

QY 1670 TTACCATCACACCCCGGAGGAGCCGAG 1698
DB 62 TTACCATCACACCCCGGAGGAGCCGAG 90

RESULT 10
AB95602
ID AB95602 standard; DNA; 1481 BP.
XX
AC AB95602;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #2100 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumor; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
FN WO200229103-A2.
XX
XX 11-APR-2002.
XX
XX 02-OCT-2001; 2001WO-US030589.
XX
XX 02-OCT-2000; 2000US-0237054P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
PI WPI; 2002-426119/45.
DR
XX
XX Diagnosing and detecting the progression of liver cancer, hepatocellular
PT carcinoma or metastatic liver tumor in a patient, involves detecting the
PT level of expression of two or more genes in a liver tissue sample.
XX
XX Claim 1; SEQ ID NO 2100; 298pp; English.

The invention relates to a novel method for diagnosing and detecting the
progression of liver cancer, hepatocellular carcinoma or metastatic liver
tumor in a patient, and differentiating metastatic liver cancer from
hepatocellular carcinoma in a patient, involving detecting the level of
expression of two or more genes represented in AB95602-AB97455 in a
tissue sample. The method of the invention has hepatotropic, and
cytostatic activity. The method is useful for diagnosing and detecting
the progression of liver cancer, hepatocellular carcinoma and metastatic
liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 1481 BP; 437 A; 381 C; 386 G; 277 T; 0 U; 0 Other;

Query Match 5.2%; Score 89; DB 6; Length 1481;
Best Local Similarity 100.0%; Pred. No. 7.1e-17;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 13:37:45 ; Search time 155 seconds
(without alignments)
6079.398 Million cell updates/sec

Title: US-09-869-508-1_COPY_453_2150

Perfect score: 1698

Sequence: 1 agaacatattgtcgcccaaa.....caccocggaggagccgcag 1698

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/PTUS_COMB.seq:*
- 6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	5.2	1481	3	US-09-429-323-3
2	83	4.9	3073	1	US-07-688-352C-31
3	83	4.9	3073	2	US-08-474-379C-31
4	83	4.9	3073	3	US-09-146-249A-31
5	83	4.9	3073	3	US-08-206-188B-31
6	83	4.9	3073	5	PCT-US91-02714-30
7	78.8	4.6	381	4	US-09-643-597-215
8	78.8	4.6	381	4	US-09-480-884A-215
9	78.8	4.6	381	4	US-09-542-615A-215
10	78.8	4.6	381	4	US-09-606-421B-215
11	76.6	4.5	7218	1	US-08-232-463-14
12	44.2	2.6	7218	1	US-08-232-463-14
13	43.4	2.6	289	3	US-09-007-005-17
14	43.4	2.6	289	3	US-09-244-796-17
15	40.8	2.4	13104	3	US-08-256-799-4
16	40.8	2.4	13104	3	US-08-462-437-4
17	37.8	2.2	832	4	US-09-621-976-2813
18	37	2.2	505	4	US-09-621-976-2813
19	36	2.1	229	4	US-09-702-705-246
20	36	2.1	229	4	US-09-736-457-246
21	36	2.1	229	4	US-09-614-124B-246
22	36	2.1	229	4	US-09-671-325-246
23	36	2.1	229	4	US-09-589-184-246
24	36	2.1	230	4	US-09-702-705-261
25	36	2.1	230	4	US-09-736-457-261
26	36	2.1	230	4	US-09-614-124B-261
27	36	2.1	230	4	US-09-671-325-261

ALIGNMENTS

RESULT 1

US-09-429-323-3
; Sequence 3, Application US/09429323A
; Patent No. 6140126
; Patent No. 6140126 6140123
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Lex M. Cowsett
; TITLE OF INVENTION: ANTISENSE MODULATION OF Y-BOX BINDING PROTEIN 1 EXPRESSION
; FILE REFERENCE: RYS-0092
; CURRENT APPLICATION NUMBER: US/09/429,323A
; CURRENT FILING DATE: 1999-10-26
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 1481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (127)..(1080)
US-09-429-323-3

Query Match

Best Local Similarity 5.2%; Score 89; DB 3; Length 1481;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1610 CGGAGCGGAGAGCGGACCCAGAGAGCCCTGAGAGCCACCGCCGCGCGGCTAG 1669
Db 2 CGGAGCGGAGAGCGGACCCAGAGAGCCCTGAGAGCCACCGCCGCGCGGCTAG 61

Qy 1670 TTACCATCACACCCCGGAGAGCGGCGAG 1698

Db 62 TTACCATCACACCCCGGAGAGCGGCGAG 90

RESULT 2

US-07-688-352C-31
; Sequence 31, Application US/07688352C
; Patent No. 5527896
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/688,352C
FILING DATE: 19910419
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael P.
REGISTRATION NUMBER: 25447
REFERENCE/DOCKET NUMBER: 27805/30197
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 3073 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..1111
US-07-688-352C-31

Query Match 4.9%; Score 83; DB 1; Length 3073;
Best Local Similarity 100.0%; Pred. No. 4.1e-16;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1616 CGAGAGCGGACCCCGAGGAGCCCTGAGCAGCCCGCCGCGCCGCTAGTTACCA 1675
DB 19 CGAGAGCGGACCCCGAGGAGCCCTGAGCAGCCCGCCGCGCCGCTAGTTACCA 78

QY 1676 TCACACCCCGGAGGAGCCCGAG 1698
DB 79 TCACACCCCGGAGGAGCCCGAG 101

RESULT 3
US-08-474-379C-31
Sequence 31, Application US/08474379C
Patent No. 5977305
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
TITLE OF INVENTION: PROCESSES
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive/6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,379C

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/688,352C
FILING DATE: 19910419
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael P.
REGISTRATION NUMBER: 25447
REFERENCE/DOCKET NUMBER: 27805/30197
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 3073 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..1111
US-07-688-352C-31

Query Match 4.9%; Score 83; DB 1; Length 3073;
Best Local Similarity 100.0%; Pred. No. 4.1e-16;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1616 CGAGAGCGGACCCCGAGGAGCCCTGAGCAGCCCGCCGCGCCGCTAGTTACCA 1675
DB 19 CGAGAGCGGACCCCGAGGAGCCCTGAGCAGCCCGCCGCGCCGCTAGTTACCA 78

QY 1676 TCACACCCCGGAGGAGCCCGAG 1698
DB 79 TCACACCCCGGAGGAGCCCGAG 101

RESULT 4
US-09-146-249A-31
Sequence 31, Application US/09146249A
Patent No. 6069240
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/146,249A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:

```

; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3073 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..1111
US-09-146-249A-31

Query Match
Best Local Similarity 100.0%; Pred. No. 4.1e-16; Length 3073;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1616 CGGAGAGCGGACCCCGAGAGCGCTGAGCAGCCCGGCGGCGGCTAGTTACCA 1675
Db 19 CGGAGAGCGGACCCCGAGAGCGCTGAGCAGCCCGGCGGCGGCTAGTTACCA 78

QY 1676 TCACACCCCGGAGAGCGCGCAG 1698
Db 79 TCACACCCCGGAGAGCGCGCAG 101

RESULT 5
US-08-206-188B-31
; Sequence 31, Application US/08206188B
; Patent No. 6100025
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; NUMBER OF SEQUENCES: 84
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/206,188B
; FILING DATE: 01-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3073 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

QY 1616 CGGAGAGCGGACCCCGAGAGCGCTGAGCAGCCCGGCGGCGGCTAGTTACCA 1675
Db 19 CGGAGAGCGGACCCCGAGAGCGCTGAGCAGCCCGGCGGCGGCTAGTTACCA 78

QY 1676 TCACACCCCGGAGAGCGCGCAG 1698
Db 79 TCACACCCCGGAGAGCGCGCAG 101

RESULT 6
PCT-US91-02714-30
; Sequence 30, Application PC/TUS9102714
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/02714
; FILING DATE: 19910419
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Borun, Michael F.
; REGISTRATION NUMBER: 25447
; REFERENCE/DOCKET NUMBER: 27805/30197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3073 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..1111
PCT-US91-02714-30

Query Match
Best Local Similarity 100.0%; Pred. No. 4.1e-16; Length 3073;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1616 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCAGCCCGCGGCGCTAGTTACCA 1675
Db 19 CGGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCAGCCCGCGGCGCTAGTTACCA 78
QY 1676 TCACACCCCGGAGGAGCGCCGAG 1698
Db 79 TCACACCCCGGAGGAGCGCCGAG 101

RESULT 7

US-09-643-597-215
; Sequence 215, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy R.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (17)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (20)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (60)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (61)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (365)
; OTHER INFORMATION: n=A,T,C or G
US-09-643-597-215

Query Match 4.6%; Score 78.8; DB 4; Length 381;

Best Local Similarity 93.0%; Pred. No. 2.3e-15;

Matches 80; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1613 GAGCGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCAGCCCGCGGCGCTAGTTA 1672
Db 1 GAGCGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCAGCCCGCGGCGCTAGTTN 60

QY 1673 CCATCACACCCCGGAGGAGCCGAG 1698

Db 61 NCATCACACCCCGGAGGAGCCGAG 86

RESULT 8

US-09-480-884A-215
; Sequence 215, Application US/09480884A
; Patent No. 6482597
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong

; APPLICANT: Fan, Liqun
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C6
; CURRENT APPLICATION NUMBER: US/09/480,884A
; CURRENT FILING DATE: 2001-08-27
; NUMBER OF SEQ ID NOS: 330
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(381)
; OTHER INFORMATION: n=A,T,C or G
US-09-480-884A-215

Query Match 4.6%; Score 78.8; DB 4; Length 381;

Best Local Similarity 93.0%; Pred. No. 2.3e-15;

Matches 80; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1613 GAGCGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCAGCCCGCGGCGCTAGTTA 1672
Db 1 GAGCGAGAGCGGACCCAGAGAGCCCTGAGCAGCCCGCCAGCCCGCGGCGCTAGTTN 60

QY 1673 CCATCACACCCCGGAGGAGCCGAG 1698

Db 61 NCATCACACCCCGGAGGAGCCGAG 86

RESULT 9

US-09-542-615A-215
; Sequence 215, Application US/09542615A
; Patent No. 6518256
; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C8
; CURRENT APPLICATION NUMBER: US/09/542,615A
; CURRENT FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 350
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (17)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (20)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (60)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (61)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (365)
; OTHER INFORMATION: n=A,T,C or G
US-09-542-615A-215

QY 510 TCTCCCTGCTCCACCAATACTCTGCTGTTTTCAGCTCTGAGTATTACCGNG 569
Db 1348 YTT
QY 570 TTGGCTGTTTAAATTTCTGCTCCATC 596
Db 1408 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1434

RESULT 12
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 2.6%; Score 44.2; DB 1; Length 7218;
Best Local Similarity 3.8%; Pred.No. 0.0034;
Matches 13; Conservative 192; Mismatches 140; Indels 0; Gaps 0;
QY 597 AGAAGGAGAACTGACTCGGGAATTTCCATCCCGAGCGATAGTAGAGCTTAAAAA 656
Db 1391 RR 1332
QY 657 AGAAGGAGAACTGAGTGGGAGGACTTCAATCAATCAGTGGCGAGCTCAATTTAT 716
Db 1331 RR 1272
QY 717 CGTTTGTGAACGCTGAGTAGTAATCCCTCTATCAGTGGCTGTGTCAGGAATAAGTGA 776

Db 1271 RR 1212
QY 777 AAAACAAACAGCTACTGTTCAATAAATGTGAGTTGAATTAATCTGATTGTGTC 836
Db 1211 RR 1152
QY 837 AGTAGAAAAAGATGTAATACTTGGAAGGAAGACACATTTTAAATATATGCTGGT 896
Db 1151 RR 1092
QY 897 AAAAGGATCAGAAGCAGGTCCCATGGAGACACACCTCGCCCT 941
Db 1091 RR 1047

RESULT 13
US-09-007-005-17/c
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rih
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ IDS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 2.6%; Score 43.4; DB 3; Length 289;
Best Local Similarity 8.2%; Pred.No. 0.0067;
Matches 20; Conservative 104; Mismatches 119; Indels 0; Gaps 0;
QY 223 TGCATTAGCTCTTAAGTCTTCTTCTACTTCCATCTCTGCTCCCTCTGCAACCATCT 282
Db 243 YGTYTAYCYGCVAYGCTYGYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNN 184
QY 283 CCACCTCCGAGCCATTTTAAAGATGCCCTCCCTACTTATGACTCTAAATGCTCT 342
Db 183 YSYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNN 124
QY 343 TCTCACTCTCCCTCAGGATATATTTCCAATTAATACTAGTCACTGCCACCTC 402
Db 123 YSYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNN 64
QY 403 TCAACCAATGTCATTCAGTCTTACTGACTTCTGACTGCTGATTCCTCCGAGATCTC 462
Db 63 YSYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNN 4
QY 463 ACC 465
Db 3 YCY 1

RESULT 14
US-09-244-796-17/c


```

; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/256,799
; FILING DATE: 06-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 88/92
; FILING DATE: 23-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: HANSSON=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: ps459
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..8834
; FEATURE:
; NAME/KEY: intron
; LOCATION: 8868..10014
; FEATURE:
; NAME/KEY: intron
; LOCATION: 10511..12277
; FEATURE:
; NAME/KEY: exon
; LOCATION: 8835..8867
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10015..10510
; FEATURE:
; NAME/KEY: exon
; LOCATION: 12278..12443
; US-08-256-799-4

Query Match 2.4%; Score 40.8; DB 3; Length 13104;
Best Local Similarity 71.1%; Pred. No. 0.065;
Matches 54; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 705 CCTCAATTTTATCGTTTGTGAAAGCTGATAGTAATCCCTCATCAGTGCCTGTTGAG 764
DB 4531 CCTCAGTTTCATCTCTCTAGATGAGGATACATAATCCCTTATGGAGTGTGAAA 4472
QY 765 GAATAAAGTGAAAAA 780
DB 4471 GGATGAAGTGAAAAA 4456

Search completed: March 28, 2004, 17:09:07
Job time : 159 secs

```


Db73AGTTACCATCACACCCCGGAGGCGCGCAG103

RESULT 2

US-09-854-133-605

Sequence 605, Application US/09854133

Publication No. US20020183499A1

GENERAL INFORMATION:

APPLICANT: Lodes, Michael J.

APPLICANT: Mchamath, Raodoh

APPLICANT: Henderson, Robert A.

APPLICANT: Benson, Darin R.

APPLICANT: Secrist, Heather

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.475C10

CURRENT APPLICATION NUMBER: US/09/854,133

CURRENT FILING DATE: 2001-05-11

NUMBER OF SEQ ID NOS: 735

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 605

LENGTH: 395

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: 375

OTHER INFORMATION: n = A,T,C or G

US-09-854-133-605

Query Match5.3%; Score 89.4; DB 9; Length 395;

Best Local Similarity98.9%; Pred. No. 8.7e-18;

Mismatches0; Conservative0; Indels0; Gaps0;

Matches90;

QY1608AGCGGAGCGGAGCGGACCCCGAGAGCCCTTGAGCAGCCCGCGCGGCT1667

Db13AGCGGAGCGGAGCGGACCCCGAGAGCCCTTGAGCAGCCCGCGCGGCT72

QY1668AGTTACCATCACACCCCGGAGGCGCGCAG1698

Db73AGTTACCATCACACCCCGGAGGCGCGCAG103

RESULT 4

US-09-880-107-2100

Sequence 2100, Application US/09880107

Patent No. US20020142981A1

GENERAL INFORMATION:

APPLICANT: Horne, Darci T.

APPLICANT: Vockley, Joseph G.

APPLICANT: Scherf, Uwe

APPLICANT: Gene Logic, Inc.

TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

FILE REFERENCE: 44921-5028-WO

CURRENT APPLICATION NUMBER: US/09/880,107

CURRENT FILING DATE: 2001-06-14

PRIOR APPLICATION NUMBER: US 60/211,379

PRIOR FILING DATE: 2000-06-14

PRIOR APPLICATION NUMBER: US 60/237,054

PRIOR FILING DATE: 2000-10-02

NUMBER OF SEQ ID NOS: 3950

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2100

LENGTH: 1481

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: Genbank Accession No. US20020142981A1 J03827

US-09-880-107-2100

Query Match5.2%; Score 89; DB 9; Length 1481;

Best Local Similarity100.0%; Pred. No. 3e-17;

Mismatches0; Conservative0; Indels0; Gaps0;

Matches89;

QY1610CGGGAGCGGAGCGGACCCCGAGAGCCCTTGAGCAGCCCGCGCGGCTAG1669

Db2CGGGAGCGGAGCGGACCCCGAGAGCCCTTGAGCAGCCCGCGCGGCTAG61

QY1670TTACCATCACACCCCGGAGGCGCGCAG1698

Db62TTACCATCACACCCCGGAGGCGCGCAG90

RESULT 5

US-10-023-969-2

Sequence 2, Application US/10023969

Publication No. US2003009598A1

GENERAL INFORMATION:

APPLICANT: Geron Corporation

APPLICANT: Irving, John

APPLICANT: Lebkowski, Jane

TITLE OF INVENTION: Chimeric Cytolytic Viruses for Cancer Treatment

FILE REFERENCE: 084,002

CURRENT APPLICATION NUMBER: US/10/023,969

CURRENT FILING DATE: 2001-12-17

PRIOR APPLICATION NUMBER: 60/256,418

PRIOR FILING DATE: 2000-12-18

NUMBER OF SEQ ID NOS: 4

SOFTWARE: PatentIn version 3.1

SEQ ID NO 2

LENGTH: 1481

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (127)..(1080)

US-10-023-969-2

Query Match5.3%; Score 89.4; DB 14; Length 395;

Best Local Similarity98.9%; Pred. No. 8.7e-18;

Mismatches0; Conservative0; Indels0; Gaps0;

Matches90;

[illegible]

Db 1 GAGCGAGAGCGGACCGGTGAGAGCCCTGAGCAGCCCCACCGCGCCGCGCTAGTTN 60
QY 1673 CCATCACACCCCGGAGGAGCGCGAG 1698
Db 61 NCATCACACCCCGGAGGAGCGCGAG 86

RESULT 13

US-09-850-716A-215
; Sequence 215, Application US/09850716A
; Patent No. US20020115139A1
; GENERAL INFORMATION:
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Retter, Marc W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C15
; CURRENT APPLICATION NUMBER: US/09/850,716A
; CURRENT FILING DATE: 2001-05-07
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (17)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (20)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (60)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (61)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (365)
; OTHER INFORMATION: n=A,T,C or G
US-09-850-716A-215

Query Match 4.6%; Score 78.8; DB 9; Length 381;
Best Local Similarity 93.0%; Pred. No. 2.3e-14;
Matches 80; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1613 GAGCGAGAGCGGACCCGAGAGCCCTGAGCAGCCCCACCGCGCCGCGCTAGTTA 1672
Db 1 GAGCGAGAGCGGACCGGTGAGAGCCCTGAGCAGCCCCACCGCGCCGCGCTAGTTN 60
QY 1673 CCATCACACCCCGGAGGAGCGCGAG 1698
Db 61 NCATCACACCCCGGAGGAGCGCGAG 86

RESULT 14

US-09-897-778-215
; Sequence 215, Application US/09897778
; Patent No. US20020147143A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Mainerakis, Margarita
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Henderson, Robert A.
; APPLICANT: Peckham, David W.
; APPLICANT: Fanger, Neil
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C16
; CURRENT APPLICATION NUMBER: US/09/897,778
; CURRENT FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 467
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 215
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 17, 20, 60, 61, 365
; OTHER INFORMATION: n = A,T,C or G
US-09-897-778-215

Query Match 4.6%; Score 78.8; DB 9; Length 381;
Best Local Similarity 93.0%; Pred. No. 2.3e-14;
Matches 80; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1613 GAGCGAGAGCGGACCCGAGAGCCCTGAGCAGCCCCACCGCGCCGCGCTAGTTA 1672
Db 1 GAGCGAGAGCGGACCGGTGAGAGCCCTGAGCAGCCCCACCGCGCCGCGCTAGTTN 60
QY 1673 CCATCACACCCCGGAGGAGCGCGAG 1698
Db 61 NCATCACACCCCGGAGGAGCGCGAG 86

RESULT 15

US-09-466-396A-215
; Sequence 215, Application US/09466396A
; Publication No. US20030119763A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C4
; CURRENT APPLICATION NUMBER: US/09/466,396A
; CURRENT FILING DATE: 1999-12-17
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (17)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (20)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (60)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (61)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (365)
; OTHER INFORMATION: n=A,T,C or G
US-09-466-396A-215

Query Match 4.6%; Score 78.8; DB 10; Length 381;
Best Local Similarity 93.0%; Pred. No. 2.3e-14;
Matches 80; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1613 GAGCGAGAGCGGACCCGAGAGCCCTGAGCAGCCCCACCGCGCCGCGCTAGTTA 1672
Db 1 GAGCGAGAGCGGACCGGTGAGAGCCCTGAGCAGCCCCACCGCGCCGCGCTAGTTN 60
QY 1673 CCATCACACCCCGGAGGAGCGCGAG 1698

Db 61 NCATCACCCCCGGGAGGAGCCGAG 86

Search completed: March 28, 2004, 19:01:23
Job time : 654 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 13:35:55 ; Search time 4640 Seconds
(without alignments)
10928.013 Million cell updates/sec

Title: US-09-869-508-1_COPY_453_2150

Perfect score: 1698

Sequence: 1 agaaccatattgtcgccaa.....caccocggaggagccgcag 1698

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hcc.*
9: gb_est1.*
10: gb_est2.*
11: gb_hcc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estom.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_nam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rtd.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gsl.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	519.4	30.6	532	9	AI659243 tui8b06.x
C 2	489.4	28.8	503	12	BI830189 603072855
C 3	451.6	26.6	565	10	BF111618 7132c03.x
C 4	451.6	26.6	608	13	EX114760 BX114760

C	5	451.6	26.6	745	14	CB850550
C	6	436	25.7	509	9	AA463628
C	7	425.4	25.1	438	9	AI650982
C	8	418.4	24.6	431	9	AI268811
C	9	410.4	24.2	428	9	AI467820
C	10	405.8	23.9	428	9	AI208001
C	11	399.4	23.5	412	14	N67996
C	12	354.4	20.9	371	10	AM448961
C	13	354.2	20.9	378	10	BE585954
C	14	352.8	20.8	866	10	BF570594
C	15	332	19.6	434	12	BI288640
C	16	327.2	19.3	543	29	CE384034
C	17	217.6	12.8	238	9	AI640509
C	18	190	11.2	803	28	AO746664
C	19	187.4	11.0	333	28	BZ867318
C	20	186.8	11.0	1016	11	AK007113
C	21	186.2	11.0	977	13	BY707189
C	22	164	9.7	703	14	CB467717
C	23	140.6	8.3	279	12	BI282773
C	24	135	8.0	600	10	BB615022
C	25	133.4	7.9	217	12	BI550827
C	26	128	7.5	913	13	EX422060
C	27	124.8	7.3	1000	13	EX450069
C	28	116	6.8	508	9	AUI26092
C	29	116	6.8	1136	12	BM459919
C	30	114.4	6.7	610	12	BI830400
C	31	112	6.6	403	12	BM853312
C	32	112	6.6	476	12	BM742560
C	33	111.4	6.6	1219	12	BM477621
C	34	110	6.5	643	10	AW163133
C	35	108	6.4	925	13	BQ421211
C	36	108	6.4	950	12	EG387948
C	37	108	6.4	1228	10	BS436330
C	38	106	6.2	950	13	BQ214006
C	39	106	6.2	1403	10	BF339614
C	40	105.4	6.2	715	12	EG773027
C	41	104.4	6.1	762	9	AI878912
C	42	103	6.1	165	12	BM847680
C	43	103	6.1	166	12	BM739099
C	44	103	6.1	358	12	BM851331
C	45	103	6.1	360	12	BM850156

ALIGNMENTS

RESULT 1
AI659243/c
LOCUS tui8b06.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251379.3',
DEFINITION mRNA sequence.
ACCESSION AI659243
VERSION AI659243.1 GI:4762813
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 532)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 463.
Location/Qualifiers
1..532

FEATURES

Source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2251379"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="NCI CGAP Pr28"
/note="Organ: Prostate; Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI CGAP Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 30.6%; Score 519.4; DB 9; Length 532;
Best Local Similarity 99.8%; Pred. No. 5.9e-137;
Matches 531; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 245 TCCTACTTCCATTCCTTGGCCCTCTGCAACCAATCTCCATCTCGGAGCCATTTTAA 304
DB 532 TCCTACTTCCATTCCTTGGCCCTCTGCAACCAATCTCCATCTCGGAGCCATTTTAA 473
QY 305 AAGATCCCTCCCTACTTATGACTCTAAATGTCTTCTCACTTCCCTCAGGATA 364
DB 472 AAGATCCCTCCCTACTTATGACTCTAAATGTCTTCTCACTTCCCTCAGGATA 413
QY 365 TATTCGAATTAATATACCTAAGTACTGCCACCTCTGCAACCAATGTCACTTGA 424
DB 412 TATTCGAATTAATATACCTAAGTACTGCCACCTCTGCAACCAATGTCACTTGA 353
QY 425 GTCTTACTGACTTACTGACTTCCGAGATCTCACTTCTTCCCTGTACCTGT 484
DB 352 GTCTTACTGACTTACTGACTTCCGAGATCTCACTTCTTCCCTGTACCTGT 293
QY 485 TGGCGGAAAGTCAGGCCCTCCACTTCTCCCTCTCCACTCCCAAAATCTTGTGT 544
DB 292 TGGCGGAAAGTCAGGCCCTCCACTTCTCCCTCTCCACTCCCAAAATCTTGTGT 233
QY 545 TTGCAGCTCTGGAGTATTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 604
DB 232 TTGCAGCTCTGGAGTATTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 173
QY 605 GAACTGACTCGGACTATTCCATCCCGCCGATAGTAGCGCTTAAAGAGACGA 564
DB 172 GAACTGACTCGGACTATTCCATCCCGCCGATAGTAGCGCTTAAAGAGACGA 113
QY 665 AGAAGTGGGTGGAGGACTTCACTAGTACAGTGGGAGCTCAATTTTATCGTTGTG 724
DB 112 AGAAGTGGGTGGAGGACTTCACTAGTACAGTGGGAGCTCAATTTTATCGTTGTG 54
QY 725 AAACGTGGATAGTAATCCCTCTATCAGTGGCTGTGTCAGGATATAAGTGAA 777
DB 53 AAACGTGGATAGTAATCCCTCTATCAGTGGCTGTGTCAGGATATAAGTGAA 1

RESULT 2

BI830189/c
LOCUS 503 bp mRNA linear EST 04-OCT-2001
DEFINITION 603072855F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5164981 5',
mRNA sequence.
BI830189
ACCESSION BI830189.1 GI:15941739
VERSION
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 503)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs@femail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Inyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11409 row: m column: 14
High quality sequence stop: 499.
Location/Qualifiers

REFERENCE

AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES

Source

1..503
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5164981"
/tissue_type="medulla"
/lab_host="DH10B"
/clone_lib="NIH_MGC_119"
/note="Organ: Brain; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: EcoRV (destroyed); RNA source normal medulla from
anonymous male age 27. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.3 kb, insert size range
0.9-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 013. Note:
this is a NIH_MGC Library."

ORIGIN

Query Match 28.8%; Score 489.4; DB 12; Length 503;
Best Local Similarity 99.6%; Pred. No. 2.2e-128;
Matches 501; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 698 GTGCAGCCTCAATTTTATCGTTTGTGAACGTTAGTAGTAATCCCTCTATCAGTGGCT 757
DB 502 GTGCAGCCTCAATTTTATCGTTTGTGAACGTTAGTAGTAATCCCTCTATCAGTGGCT 443
QY 758 GTTCAGAGATAAGTCAAAAAACAAACAGGCTAGCTTCTCAATAAATGTGAGTTGAA 817
DB 442 GTTCAGAGATAAGTCAAAAAACAAACAGGCTAGCTTCTCAATAAATGTGAGTTGAA 384
QY 818 TTAATCTGATTTGTGCTAGTAGAATAAGATGTGAATCTTGGAAAGGAGACACATTT 877
DB 383 TTAATCTGATTTGTGCTAGTAGAATAAGATGTGAATCTTGGAAAGGAGACACATTT 324
QY 878 TTTTAAATATATGCTCGTAAACGATCAGAGGAGGTCCTCCATGAGACACACCTCG 937
DB 323 TTTTAAATATATGCTCGTAAACGATCAGAGGAGGTCCTCCATGAGACACACCTCG 264
QY 938 CCTTAAATATGCTGAAACCGGCTGCTAGCTTCTCAATAAATGTGAGTTGAA 997
DB 263 CCTTAAATATGCTGAAACCGGCTGCTAGCTTCTCAATAAATGTGAGTTGAA 204
QY 998 CGACAAAAGGTCAGCTTCTTCAACGCTATGTTTAAAGGCAATTCAGAAACCTCGGCT 1057
DB 203 CGACAAAAGGTCAGCTTCTTCAACGCTATGTTTAAAGGCAATTCAGAAACCTCGGCT 144
QY 1058 GTGCGGAGCTACACGCGCATTTAAAGAAAAGACGACTTATGCCCGCGTAAATGTTCTCA 1117
DB 143 GTGCGGAGCTACACGCGCATTTAAAGAAAAGACGACTTATGCCCGCGTAAATGTTCTCA 84
QY 1118 GATCACAGGACCGTATTGTGAGCTGGAGGAGGAGGAGCGCTTTTCTTACGGGGGGCTA 1177
DB 83 GATCACAGGACCGTATTGTGAGCTGGAGGAGGAGGAGCGCTTTTCTTCTACGGGGGGCTA 24

```
QY 1178 AGGCGTCTTCGAGCCCTTCCA 1200
|||||
Db 23 AGGCGTCTTCGAGCCCTTCCA 1

RESULT 3
BF111618/c
LOCUS BF111618.1 565 bp mRNA linear EST 20-OCT-2000
DEFINITION IMAGE:3523180.3', mRNA sequence.
ACCESSION BF111618
VERSION BF111618.1 GI:10941308
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 565)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 463.
FEATURES
source
1. 565
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3523180"
/lab_host="DH10B"
/clone_lib="Soares NSF P8 9W OT PA P S1"
a note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHSP pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NBHFA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."

Query Match 26.6%; Score 451.6; DB 10; Length 565;
Best Local Similarity 98.9%; Pred. No. 1.5e-117;
Matches 465; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 353 CCCCTCAGGATATATTCCAATTAATACTAAGTGAAGTCCACCTCTGCAACCCAA 412
|||||
Db 472 CACTTGAGGATATATTCCAATTAATACTAAGTGAAGTCCACCTCTGCAACCCAA 413
|||||
QY 413 TGTACATTCAGTCTTACTCACTTGAAGTCTGATTCGATTCCTCCGAGATCTCACTCTTCTC 472
|||||
Db 412 TGTACATTCAGTCTTACTCACTTGAAGTCTGATTCGATTCCTCCGAGATCTCACTCTTCTC 353
|||||
QY 473 GCCTGTACCTGTGCGGGAAGTCAGCCCTCCACCTTCTCCCTGCTTCACTCCCAAAA 532
|||||
Db 352 GCCTGTACCTGTGCGGGAAGTCAGCCCTCCACCTTCTCCCTGCTTCACTCCCAAAA 293
|||||
QY 533 TACTTCGGTGTTCAGCTCTGAGATATTTACCGTGTGGCTGTAAATTTCTGCCTC 592
|||||
```

```
Db 292 TACTTCGTGTTTTTCAGCTCTGGAGTATTACCGTGTGGCTGTAAATTTCTGCCTC 233
|||||
QY 593 CATCAGAGGCGAGAACTGACTCGGGAACCTATTTCATCCCGAGCCGATAGTAGCGCTTA 652
|||||
Db 232 CATCAGAGGCGAGAACTGACTCGGGAACCTATTTCATCCCGAGCCGATAGTAGCGCTTA 173
|||||
QY 653 AAAAGAACGGAAGAGGTGGTGGGAGGACTTCAGTAACATCAGGTGGCAGCCCTCAATT 712
|||||
Db 172 AAAAGAACGGAAGAGGTGGTGGTGGGACTTCAGTAACATCAGGTGGCAGCCCTCAATT 114
|||||
QY 713 TTATCGTTTGTGAACGCTGAGTAGTAATCCCTCTATCAGCTGGCTGTGTCAGGAATAAAG 772
|||||
Db 113 TTATCGTTTGTGAACGCTGAGTAGTAATCCCTCTATCAGCTGGCTGTGTCAGGAATAAAG 54
|||||
QY 773 TGAATAAACCAACAGGCTAGCTTCTCAATAAATGTGAGTTGAATTA 822
|||||
Db 53 TGAATAAACCAACAGGCTAGCTTCTCAATAAATGTGAGTTGAATTA 4

RESULT 4
LOCUS BX114760
DEFINITION BX114760 Soares NHMPu S1 Homo sapiens cDNA clone IMAGE:811834 ;
IMAGE:811834, mRNA sequence.
ACCESSION BX114760
VERSION BX114760.1 GI:27879797
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 608)
Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,
Radelof, U., Schneider, D. and Korn, B.
Human Unigeneset - RZPD3
Unpublished (2003)
Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGE:811834
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Human Unigeneset - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi?response?libNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTCACAGGAACAGCTATGAC.
FEATURES
Location/Qualifiers
1..608
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:811834"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/clone_lib="Soares NHMPu S1"
a note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not 1;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
```

```
ORIGIN
Query Match      26.6%; Score 451.6; DB 13; Length 608;
Best Local Similarity 98.9%; Pred. No. 1.6e-117;
Matches 465; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 353 CCCTCAGGATATATTTCCAAATTAATATACCTTAAGTACCTGCCACCTCTGCAACCCAA 412
DB 128 CACTTGAGGATATATTTCCAAATTAATATACCTTAAGTACCTGCCACCTCTGCAACCCAA 187
QY 413 TGTCAATTCGAGTCTTACTGAACACTTCTGACATGCAATTTCCCGAGATCTCACTCTTCTC 472
DB 188 TGTCAATTCGAGTCTTACTGAACACTTCTGACATGCAATTTCCCGAGATCTCACTCTTCTC 247
QY 473 GCCTGTACCTCTGCGGGAAAGTCAGCCCTCCACCTTCTCCTGCTTCCACTCCCAAA 532
DB 248 GCCTGTACCTCTGCGGGAAAGTCAGCCCTCCACCTTCTCCTGCTTCCACTCCCAAA 307
QY 533 TACTTCGTGGTTTTCGAGCTCTGAGATTAATACCGTGTGGCTGTGTTAAATTTCTGCCCTC 592
DB 308 TACTTCGTGGTTTTCGAGCTCTGAGATTAATACCGTGTGGCTGTGTTAAATTTCTGCCCTC 367
QY 593 CATCAGAAGCGAGAAACTGACTCGGAACATTAATCCATCCCGAGCGATAGAGCCTTA 652
DB 368 CATCAGAAGCGAGAAACTGACTCGGAACATTAATCCATCCCGAGCGATAGAGCCTTA 427
QY 653 AAAAAGAACGGAAGAGTGGGTGGAGGACTTCAGTAACATCAGTGGCAGCTCAATT 712
DB 428 AAAAAGAACGGAAGAGTGGGTGGAGGACTTCAGTAACATCAGTGGCAGCTCAATT 486
QY 713 TTATCGTTTGTGAACCGTGATAGTAATCCCTCTATCAGTGGCTGTGCGAGAAATAAG 772
DB 487 TTATCGTTTGTGAACCGTGATAGTAATCCCTCTATCAGTGGCTGTGCGAGAAATAAG 546
QY 773 TGAATAAACCAAAACAGCGCTAGCTGTGTTCAATAAATGTGAGTTCAATTAA 822
DB 547 TGAATAAACCAAAACAGCGCTAGCTGTGTTCAATAAATGTGAGTTCAATTAA 596

RESULT 5
CB850550/c
LOCUS
DEFINITION
UI-CP-EN1-adc-h-05-0-UI.s1 UI-CP-EN1 Homo sapiens cDNA clone
UI-CP-EN1-adc-h-05-0-UI.s1, mRNA sequence.
CB850550
CB850550.1 GI:30045317
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 745)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
PUBMED
8889548
COMMENT
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
```

```
sequence: 1-22, >AT_rich#Low_complexity 542-620, >LINE2 (matched
compliment)
Seq primer: M13 FORWARD
POLYA=Yes.
FEATURES
Location/Qualifiers
1..745
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CP-EN1-adc-h-05-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (TI phage resistant)"
/clone_lib="UI-CP-EN1"
/note="Torsan; Lung; Vector: p7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: BcoR I; Site 2: Not I;
UI-CP-EN1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Cystic Fibrosis
Epithelial Cells. The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an BcoR I adaptor, digested with Not
I, and cloned directionally into p7T3-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CTGCTCAGGT.
TAG_SEQ=None found"
ORIGIN
Query Match      26.6%; Score 451.6; DB 14; Length 745;
Best Local Similarity 98.9%; Pred. No. 1.7e-117;
Matches 465; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 353 CCCCTCAGGATATATTTCCAAATTAATATACCTTAAGTACCTGCCACCTCTGCAACCCAA 412
DB 484 CACTTGAGGATATATTTCCAAATTAATATACCTTAAGTACCTGCCACCTCTGCAACCCAA 425
QY 413 TGTCAATTCGAGTCTTACTGAACACTTCTGACATGCAATTTCCCGAGATCTCACTCTTCTC 472
DB 424 TGTCAATTCGAGTCTTACTGAACACTTCTGACATGCAATTTCCCGAGATCTCACTCTTCTC 365
QY 473 GCCTGTACCTCTGCGGGAAAGTCAGCCCTCCACCTTCTCCTGCTTCCACTCCCAAA 532
DB 364 GCCTGTACCTCTGCGGGAAAGTCAGCCCTCCACCTTCTCCTGCTTCCACTCCCAAA 305
QY 533 TACTTCGTGGTTTTCGAGCTCTGAGATTAATACCGTGTGGCTGTGTTAAATTTCTGCCCTC 592
DB 304 TACTTCGTGGTTTTCGAGCTCTGAGATTAATACCGTGTGGCTGTGTTAAATTTCTGCCCTC 245
QY 593 CATCAGAAGCGAGAAACTGACTCGGAACATTAATCCATCCCGAGCGATAGAGCCTTA 652
DB 244 CATCAGAAGCGAGAAACTGACTCGGAACATTAATCCATCCCGAGCGATAGAGCCTTA 185
QY 653 AAAAAGAACGGAAGAGTGGGTGGAGGACTTCAGTAACATCAGTGGCAGCTCAATT 712
DB 184 AAAAAGAACGGAAGAGTGGGTGGAGGACTTCAGTAACATCAGTGGCAGCTCAATT 126
QY 713 TTATCGTTTGTGAACCGTGATAGTAATCCCTCTATCAGTGGCTGTGCGAGAAATAAG 772
DB 125 TTATCGTTTGTGAACCGTGATAGTAATCCCTCTATCAGTGGCTGTGCGAGAAATAAG 66
QY 773 TGAATAAACCAAAACAGCGCTAGCTGTGTTCAATAAATGTGAGTTCAATTAA 822
DB 65 TGAATAAACCAAAACAGCGCTAGCTGTGTTCAATAAATGTGAGTTCAATTAA 16

RESULT 6
AA463628/c
LOCUS
DEFINITION
AA463628 Soares_NHNPu_S1 Homo sapiens cDNA clone IMAGE:811834
zx98906.s1 Soares_NHNPu_S1 Homo sapiens cDNA clone IMAGE:811834
```

3', mRNA sequence.
AA463628
AA463628.1 GI:2188512
EST.
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,
Moore, B., Schellenger, K., Steptoe, W., Tan, F., Theising, J.,
White, Y., Wyllie, T., Waterston, R., and Wilson, R.
WashU-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 486.
Location/Qualifiers
1. 509
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:6042745"
/db_xref="taxon:9606"
/clone="IMAGE:811834"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/clone_lib="Soares NHMPU S1"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not 1;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
FEATURES
source
1. 509
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:6042745"
/db_xref="taxon:9606"
/clone="IMAGE:811834"
/tissue_type="Pooled germ cell tumors"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs:
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."
ORIGIN
Query Match 25.7%; Score 436; DB 9; Length 509;
Best Local Similarity 98.5%; Pred. No. 4.3e-113;
Matches 461; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
QY 353 CCCCTCAGGATATATTTCCCAATTAATAATACCTAAGTGACTGCCACCTCTGCAACCCAA 412
DB 467 CACTGTGAGATATATTTCCCAATTAATAATACCTAAGTGACTGCCACCTCTGCAACCCAA 408
QY 413 TGTCACTTCGAGTCTTACTGAACTACTTGACTGCATTTCCGAGATCTCACTCTTCTC 472
DB 407 TGTCACTTCGAGTCTTACTGAACTACTTGACTGCATTTCCGAGATCTCACTCTTCTC 348
QY 473 GCCTGTACCTGTGCGGGGAAGTCAGCCCTTCCACCTTCTCCCTGCTTCCACTCCC-AAA 531
DB 347 GCCTGTACCTGTGCGGGGAAGTCAGCCCTTCCACCTTCTCCCTGCTTCCACTCCCAGAA 288
QY 532 ATACTTGTGTTTTCGAGCTCTGGAGTATTACCTGTTGGCTGTTAAATTTCTGCCT 591
DB 287 ATACTTGTGTTTTCGAGCTCTGGAGTATTACCTGTTGGCTGTTAAATTTCTGCCT 228
QY 592 CCATCAGAGCGAAGAACTGACTCGCAACTATTCCATCCCGACCGATAGTAGAGCGCTT 651
DB 227 CCATCAGAGCGAAGAACTGACTCGCAACTATTCCATCCCGACCGATAGTAGAGCGCTT 168

652 AAAAAAGAACGGAAGAGGTGGTGGAGGACTTCAGTAACATCAGGTGGCGAGCTCAAT 711
167 AAAAAAGAACGGAAGAGGTGGT- GGAGGACTTCAGTAACATCAGGTGGCGAGCTCAAT 109
712 TTTATCGTTTGTGAACGTTGATAGTATCCCTCTATCACGTGGCTGTTCGAGGAATAAA 771
108 TTTATCGTTTGTGAACGTTGATAGTATCCCTCTATCACGTGGCTGTTCGAGGAATAAA 49
772 GTGAAAAACAAACACAGGCTAGCTTGTTCATATAATGTGAGTTGAATT 819
48 GTGAAAAACAAACACAGGCTAGCTTGTTCATATAATGTGAGTTGAATT 1
RESULT 7
LOCUS
AI650982 438 bp mRNA linear EST 16-DEC-1999
DEFINITION
wa96d06.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2304011 3',
mRNA sequence.
ACCESSION
AI650982
VERSION
AI650982.1 GI:4734961
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. (bases 1 to 438)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 552 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. 438
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2304011"
/tissue_type="Pooled germ cell tumors"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_GC6"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs:
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."
ORIGIN
Query Match 25.1%; Score 425.4; DB 9; Length 438;
Best Local Similarity 99.8%; Pred. No. 4.4e-110;
Matches 426; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 858 TTGGAAGGAGACACACATTTTAAATATATATGCTGTGTAACCGATCAGAAGCAGGT 917
DB 3 TTGGAAGGAGACACACATTTTAAATATATATGCTGTGTAACCGATCAGAAGCAGGT 62
QY 918 CCCCTGAGAGCACCCCTGCGCTTAAACATGCTGAACCCGGGCTGCCATAGCTCGTGG 977

Db 63 CCCCATGGAGACACCCCTGCGCTTAAACATGCTGAACCGGGCTGCCATAGCTGCGTGG 122

QY 978 TCCCTCCCAAGGTGACTGCTCCGACAAAAGGGTACCTCTTCAAAAGCATAGCTTTAAGGC 1037

Db 123 TCCCTCCCAAGGTGACTGCTCCGACAAAAGGGTACCTCTTCAAAAGCATAGCTTTAAGGC 182

QY 1038 AATTCAGAAAACCTTCGGCTGTGCGCGGACTACACGGCCATTAAGAAAAGACGACTCTA 1097

Db 183 AATTCAGAAAACCTTCGGCTGTGCGCGGACTACACGGCCATTAAGAAAAGACGACTCTA 242

QY 1098 TCCCGCCGCTAATGTTCTCAGATCACAGGACCGTATTGGAGCTGGGAGGGAGGAGC 1157

Db 243 TCCCGCCGCTAATGTTCTCAGATCACAGGACCGTATTGGAGCTGGGAGGGAGGAGC 302

QY 1158 CTTTCTTTCACGGGGGGCTTAAGCGCTTCGAGCGCCCTTCCAAATCCCGGGTCCGGCCGG 1217

Db 303 CTTTCTTTCACGGGGGGCTTAAGCGCTTCGAGCGCCCTTCCAAATCCCGGGTCCGGCCGG 362

QY 1218 GTAATCCCTGCCAGCGTTTCGGCGTGCCTTTTTCAGCCGAGACACAAACCTGAACGT 1277

Db 363 GTAATCCCTGCCAGCGTTTCGGCGTGCCTTTTTCAGCCGAGACACAAACCTGAACGT 422

QY 1278 GGGGGCC 1284

Db 423 GGGGGCC 429

RESULT 8

AI268811/c

LOCUS

DEFINITION

q040g10.x1 NCI CGAP Lu5 Homo sapiens CDNA clone IMAGE:1911042 3'

similar to contains OFR.b3 OFR repetitive element ; , mRNA sequence.

AI268811

AI268811.1 GI:3887978

EST.

Source

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 431)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 408.

Location/Qualifiers

1. 431

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:1911042"

/tissue_type="carcinoid"

/lab_host="DH10B"

/clone_lib="NCI CGAP Lu5"

/notes="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 24.6%; Score 418.4; DB 9; Length 431;

Best Local Similarity 99.5%; Pred. No. 4.4e-108;

Matches 430; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 407 ACCCAATGTCACTTTCAGTCTTACTGAACCTACTTGCATGCTTCCGAGATCTCACCT 466

Db 431 ACCCAATGTCACTTTCAGTCTTACTGAACCTACTTGCATGCTTCCGAGATCTCACCT 372

QY 467 CTTCTCGCTGTACCTGTGCGGGAAGTTCAGCCCTCCACCTTCTCCCTGCTTCCACTC 526

Db 371 CTTCTCGCTGTACCTGTGCGGGAAGTTCAGCCCTCCACCTTCTCCCTGCTTCCACTC 312

QY 527 CCAAAATACTTCTGCTGTTTTCAGCTCTCTGGAGTATTATTCGTTGGCTGTTTAAATTC 586

Db 311 CCAAAATACTTCTGCTGTTTTCAGCTCTCTGGAGTATTATTCGTTGGCTGTTTAAATTC 252

QY 587 TGCTTCATCAGAGGAGGAACTGACTCGCGAACTATTCATCCACCCAGCCGATAGTAGA 646

Db 251 TGCTTCATCAGAGGAGGAACTGACTCGCGAACTATTCATCCACCCAGCCGATAGTAGA 192

QY 647 CGCTTAAAAAGAACGGAAGGTGGTGGAGGACTTCAGTAACATCAGGTGGCAGCC 706

Db 191 CGCTTAAAAAGAACGGAAGGTGGTGGAGGACTTCAGTAACATCAGGTGGCAGCC 133

QY 707 TCAATTTTATCGTTTGTGAAACCTGATAGTAACTCCCTCTATCACGTGGCTGTTGCAGGA 766

Db 132 TCAATTTTATCGTTTGTGAAACCTGATAGTAACTCCCTCTATCACGTGGCTGTTGCAGGA 73

QY 767 ATAAAGTAAAAACAAAACAGAGCTAGCTTGTTCATTAATGAGTGAATTAATCTG 826

Db 72 ATAAAGTAAAAACAAAACAGAGCTAGCTTGTTCATTAATGAGTGAATTAATCTG 13

QY 827 ATTGTGGTCTAG 838

Db 12 ATTGTGGTCTAG 1

RESULT 9

AI467820/c

LOCUS

DEFINITION

t378b11.x1 Soares NSF P8 9W OT PA P S1 Homo sapiens CDNA clone IMAGE:2147613 3' similar to contains OFR.b3 OFR repetitive element ; , mRNA sequence.

AI467820

AI467820.1 GI:4329910

EST.

Source

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 428)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert length: 576 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 393.

Location/Qualifiers

1. 428

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2147613"

/lab_host="DH10B"

/clone_lib="Soares NSF P8 9W OT PA P S1"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with

the driver population), plus a pool of 11,136 clones from NCI-CGAP Sub3 (IMAGE Clonides 2712456-2723591) (10% of the driver population), plus a pool of 5,472 clones from NCI-CGAP Sub4 (IMAGE Clonides 2723592-2728969) (70% of the driver population). Subtraction was performed as previously described [Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery. Genome Research 6, 791-806. TAG_TISSUE=germ cell TAG_LIB=NCI CGAP_GC4 TAG_SEQ=AAATC"

ORIGIN

Query Match 20.9%; Score 354.4; DB 10; Length 371;
Best Local Similarity 99.7%; Pred. No. 8.5e-90;
Matches 355; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 858 TTGGAAGGAGACACATTTTTTAAATATATGCTGTAAACGGATCAGAGCAGGT 917
Db 16 TTGGAAGGAGACACATTTTTTAAATATATGCTGTAAACGGATCAGAGCAGGT 75
QY 918 CCCATGGAGCACACCTCGCCCTAAACATGCTGAACCGGCTGCCATGCTGCTGG 977
Db 76 CCCATGGAGCACACCTCGCCCTAAACATGCTGAACCGGCTGCCATGCTGCTGG 135
QY 978 TCCTCCAAAGTGACTGTCTCCAGCAAAAGGGTACGCTCTTCAACGCGATACGTTTAAAGC 1037
Db 136 TCCTCCAAAGTGACTGTCTCCAGCAAAAGGGTACGCTCTTCAACGCGATACGTTTAAAGC 195
QY 1038 AATTCAGAAACCTCGGCTGTGCGCGACTACACGGCCATTAAGAAAGACGACTCTA 1097
Db 196 AATTCAGAAACCTCGGCTGTGCGCGACTACACGGCCATTAAGAAAGACGACTCTA 255
QY 1098 TGGCCGCGTATGTTCTCAGATCAGCAGGACCGTATTGGAGCTGGAGGAGGAGGAGC 1157
Db 256 TGGCCGCGTATGTTCTCAGATCAGCAGGACCGTATTGGAGCTGGAGGAGGAGGAGC 315
QY 1158 CTTTCTTTCACGGGGGGTAAAGCGCTTTCGAGCCCGCTTCCAAATCCCGGGTCCGG 1213
Db 316 CTTTCTTTCACGGGGGGTAAAGCGCTTTCGAGCCCGCTTCCAAATCCCGGGTCCGG 371

RESULT 13

BE858954/c
LOCUS BE858954
DEFINITION 7944f12.x1 NCI CGAP Pr28 Homo sapiens cDNA clone IMAGE:3309359 3', mRNA sequence.
ACCESSION BE858954
VERSION BE858954.1 GI:10374531
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 378)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 342.
Location/Qualifiers 1. 378

FEATURES

source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3309359"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="NCI CGAP Pr28"
/notes="Organ: Prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; plasmid DNA from the normalized library NCI CGAP Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (Clonides 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 20.9%; Score 354.2; DB 10; Length 378;
Best Local Similarity 97.6%; Pred. No. 9.8e-90;
Matches 370; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
QY 404 GCAACCAATGTCACATTCGAGTCTTACTGAACCTACTTGACTGATCTCCCGAGATCTCA 463
Db 378 GCAACCAATGTCACATTCGAGTCTTACTGAACCTACTTGACTGATCTCCCGAGATCTCA 319
QY 464 CTTCTTCGCTGTACCTGTGCGGGAAGTACGCTTCCACCTTCCCTGCTTCCA 523
Db 318 CTTCTTCGCTGTACCTGTGCGGGAAGTACGCTTCCACCTTCCCTGCTTCCA 259
QY 524 CTCCCAAAATACCTCGTGGTTTGGAGCTCTGGAGTATTACCGTCTGGCTGTTAAAT 583
Db 258 CTCCCAAAATACCTCGTGGTTTGGAGCTCTGGAGTATTACCGTCTGGCTGTTAAAT 199
QY 584 TTCTGCTCCATCAGAGGAGCAAACTGATCCGGAATATTCCATCCCGAGCGATAGT 643
Db 198 TTCTGCTCCATCAGAGGAGCAAACTGATCCGGAATATTCCATCCCGAGCGATAGT 139
QY 644 AGACGCTTAAAAAGAACGGAAGAGTGGTGGAGGACTTCAGTAAATCAGGTGGCA 703
Db 138 AGACGCTTAAAAAGAACGGAAGAGTGGTGGAGGACTTCAGTAAATCAGGTGGCA 80
QY 704 GCCTCAATTTATCGTTTGTGAACGTTGATAGTATCCCTCTATCAGTGGCTGTGCA 763
Db 79 GCCTCAATTTATCGTTTGTGAACGTTGATAGTATCCCTCTATCAGTGGCTGTGCA 20
QY 764 GGAATAAAGTGAAAAACA 782
Db 19 GGAATAAATAAAGGCCCA 1

RESULT 14

BF570994/c
LOCUS BF570994
DEFINITION 502075915F1 NIH MGC_62 Homo sapiens cDNA clone IMAGE:4243371 5', mRNA sequence.
ACCESSION BF570994
VERSION BF570994.1 GI:11644706
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 866)
AUTHORS NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC/DC/DTP
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

1. .866

LOCATION/ QUANTITY
1. .866

Overall Match 20 88% Score 352 8 DB 10: Length 866:

701 GCAGCC¹TCAATTTTATCGTTGTGAAACGTGGATAGTAATCCC-TCTATCACGTGGCTGT 759

[illegible]

1000

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

463 TTTGTGAAATATAATGCCTTGGTTAAACGGAICAGAAAGGCAAGGTCCTCCCAIGGAGGACACCC 467

403 TCGCCCT-AACATGCTGGAACCGGGCTTGCCATACCTGCGTGGTCCCTCCACAGGTGACT 345

344 TGCTCCGGTACACAAAGGGTACGCTCTTTCAACGCATACGTTTAAGGCCATTCCAGTAAC 285

284 CCTCTGGCTGCCGGACTACAC-GCCATTAAAGAAAGCAGCTCTATGCC-CCGTAA 227

226 TGTTCTCAGATCACAGGGACCGTATTGGA - CTGGAGGGAGGGAAGCCTTTTCTTCAACC 168

167 GGGGGCTAAGGCGTCTTCGAGCCCCCTTCCATCCCCGTCCGCACGGTAATCCCTGC 108

[illegible]

RESULT 15

LOCUS
DEPTINTO

ACCESSION

KEYWORDS
SOURCE

AUTHORS

JOURNAL
COMMENT

2nos

Query M

Matches

25

8

DB

57

D'b

 ∂y

Db 241 TAGCGGAGCGGAGAGCGGACCCACAGAGAGCCCTGAGCAGCCCCACCGCCGCCGCCGCC 300
Qy 1667 TAGTTACCATCACACCCCGGAGGAGCGCGCAG 1698
Db 301 TAGTTACCATCACACCCCGGAGGAGCGCGCAG 332

Search completed: March 28, 2004, 17:06:24
Job time : 4645 secs